

**NATURALISING THE A PRIORI:
RELIABILISM AND EXPERIENCE-INDEPENDENT
KNOWLEDGE**

BENEDICT YOUNG

**PhD
EDINBURGH UNIVERSITY
2000**



This thesis is my own work. All influences have been properly cited.

Ben Young
Edinburgh University
April 2000

The research for this thesis was funded by George Bennett and Leonie Wilson; thanks to them. Special thanks also to Tim Williamson, Richard Gray and Denis Walsh.

Material from this thesis was presented at the *Mind, World and Knowledge* conference, Reading, December 1997; *Spring Conference of the SPPA*, Edinburgh 1998; *Joint Session of the Mind/Aristotelian Society*, University of Hertfordshire, 1998; 2nd and 3rd *National Postgraduate Analytic Philosophy Conferences*, Sheffield, June 1998 and 1999; *25th IUC Annual Philosophy of Science Conference*, Dubrovnik, Croatia, April 1999; *Royal Institute of Philosophy Conference*, Edinburgh, July 1999; and in Honours Undergraduate lectures on A Priori Knowledge delivered at Edinburgh University 1998. Thanks to audiences at these events, to students in my Honours epistemology classes 1999-2000, and to Alexander Bird, Bob Fogelin, Dave Hemp, Paul Horwich, and Mike Martin for useful comments and discussions.

More thanks to: Alastair, Ariel, Emily, Eleni, Hild, Katja, Lucas, Majeda, Pete, Rebecca and Sophie; the staff of DHT Language and Humanities centre; and Leonie and George again.

ABSTRACT

The thesis defends the view that the concept of a priori knowledge can be naturalised without sacrificing the core aspects of the traditional conception of apriority. I proceed by arguing for three related claims.

The first claim is that the adoption of naturalism in philosophy is not automatically inconsistent with belief in the existence of a priori knowledge. A widespread view to the contrary has come about through the joint influence of Quine and the logical empiricists. I hold that by rejecting a key assumption made by the logical empiricists (the assumption that apriority can be explained only by appeal to the concept of analyticity), we can develop an account of naturalism in philosophy which does not automatically rule out the possibility of a priori knowledge, and which retains Quine's proposals that philosophy be seen as continuous with the enterprise of natural science, and that the theory of knowledge be developed within the conceptual framework of psychology.

The first attempt to provide a theory of a priori knowledge within such a framework was made by Philip Kitcher. Kitcher's strategy involves giving an account of the idea of "experience-independence" independently of the theory of knowledge in general (he assumes that an appropriate account of the latter will be reliabilist). Later authors in the tradition Kitcher inaugurated have followed him on this, while criticising him for adopting too strong a notion of experience-independence. The second claim I make is in qualified agreement with this: it is that only a weak notion of experience-independence will give a viable account of a priori knowledge, but that the reasons why this is so have been obscured by Kitcher's segregation of the issues. Strong reasons for adopting a weak notion are provided by consideration of the theory of knowledge, but these same reasons also highlight severe problems for the project of providing a naturalistic theory of knowledge in general.

The third claim is that a plausible naturalistic theory of knowledge in general can be given, and that it provides an appropriate framework within which to give an account of minimally experience-independent knowledge.

I conclude with a consideration of some of the problems that an account of minimal a priori knowledge will have to address.

TABLE OF CONTENTS

INTRODUCTION	1
CONCEPTUAL FRAMEWORK	9
CARNAP'S ACCOUNT OF A PRIORI KNOWLEDGE	13
MODERATE EMPIRICISM: THE ANALYTIC THEORY OF A PRIORI KNOWLEDGE	14
CARNAP'S ACCOUNT OF LANGUAGE AND ANALYTICITY	18
QUINE'S "TRUTH BY CONVENTION"	24
CARNAP'S RESPONSE	30
THE PROBLEM OF LOGICAL AND PHYSICAL RULES	33
QUINE'S CRITIQUE OF ANALYTICITY	39
QUINE'S "TWO DOGMAS OF EMPIRICISM"	40
INTERPRETATION AND DISCUSSION	53
QUINE'S LEGACY	63
EVALUATING THE ANALYTIC THEORY	64
NATURALISM AND A PRIORI KNOWLEDGE	69
OVERVIEW OF THEORIES OF THE A PRIORI	69
EPISTEMOLOGICAL NATURALISM	73
METAPHYSICAL NATURALISM	80
NATURALISED EPISTEMOLOGY AND A PRIORI KNOWLEDGE	85
UNREVISABILITY	95
EXPERIENCE-INDEPENDENCE	95
THE CONCEPT OF UNREVISABILITY	97
THE UNREVISABILITY THESIS	104
KITCHER'S ATTEMPT TO MOTIVATE THE UNREVISABILITY THESIS	104
CASULLO'S ARGUMENT AGAINST THE UNREVISABILITY THESIS	108
THE ARGUMENT FROM THE REVISABILITY OF KNOWLEDGE IN GENERAL	114
INDIRECT REVISION AND IMAGINABILITY	117
DIRECT REVISION AND SCEPTICISM ABOUT A PRIORI KNOWLEDGE	119
CONCLUSION	122
INFALLIBILITY	125
JUSTIFICATION	128
WARRANT	134

GOLDMAN'S RELIABILISM	139
"A CAUSAL THEORY OF KNOWING"	141
"DISCRIMINATION AND PERCEPTUAL KNOWLEDGE"	145
DISCRIMINATION	146
RELEVANCE	149
CLAUSE 4	150
"WHAT IS JUSTIFIED BELIEF?"	153
EPISTEMOLOGY AND COGNITION	158
KNOWLEDGE	158
THE GENERALITY PROBLEM	161
DEVIANT PROCESSES	165
JUSTIFICATION	167
GOLDMAN'S 1986 ANALYSIS OF KNOWLEDGE	170
 PROCESS RELIABILISM AND CONTEXTUALISM	 173
 CLOSURE AND RELEVANCE	 175
LEWIS'S ACCOUNT OF "PROPER IGNORING"	182
CONTEXTUAL RELIABILITY	185
PROBLEMS FOR CONTEXTUALISM	193
NONREDUCTIVE NATURALISM	199
CONCLUSION	203
 NONEXPERIENTIALITY	 205
 EXPLAINING "NONEXPERIENTIAL"	 206
BOGHOSSIAN'S IMPLICIT DEFINITION ACCOUNT	213
BOGHOSSIAN'S ACCOUNT OF A PRIORI KNOWLEDGE	218
INNATE AND A PRIORI KNOWLEDGE	223
CONCLUSION	235
 BIBLIOGRAPHY	 239

CHAPTER ONE

Introduction

Consider how we would come to know the following propositions: *all humans are inquisitive*; and $5+7=12$. Assuming the first to be true, the most direct means of coming to know it would be to gather evidence about the psychology of humans. Since it would be impractical to examine each and every individual human, we would proceed by sampling the human population and making an inductive generalisation based on that sample. Whether we finally come to know that all humans are inquisitive through this method depends on the size and quality of the sample we gather. The sample must be based on observations of many humans; but it must also be drawn from a wide range of times, places and social contexts: in other words, it must be representative of the human population as a whole. It is not at all easy to say precisely what is the relationship between a sample and the inductive knowledge drawn from it, but in general terms the following must be true: the larger the sample, and the wider the variation in parameters that it sustains, the better our chances of deriving knowledge from it.

Contrast this with the typical way in which we come to know the second proposition: that is, through the use of mental arithmetic. Note first that our knowledge that five and seven make twelve is not totally insensitive to evidence from the senses. It seems reasonable to say that someone who has just counted two (disjoint) sets of five and seven bananas respectively may sometimes be in a better position to know $5+7=12$ than another who has not. Perhaps the second person slips up in his mental arithmetic; the first, using the bananas as a guide, is less likely to have made a mistake. But the counting of the two sets of bananas stands in a completely different relationship to our knowledge that $5+7=12$, than does the sampling process with respect to our knowledge that all humans are inquisitive.¹ For in the case of arithmetic, it seems unnecessary, even wrong, to continue making observations about the numerical relations between sets of objects beyond an initial check. We would not look favourably on someone who said "I have seen that five

¹ See Kim 1981.

Introduction

and seven bananas make twelve bananas, but before I commit myself to the claim that five and seven make twelve in all cases I would want to check with other objects, preferably non-bananas." In contrast, it would be reasonable to withhold judgement as to whether all humans were inquisitive after having been shown, say, that all children were inquisitive, or all dustmen were inquisitive: in this case it is entirely in order to want to see whether some further sample, preferably in appropriate respects quite different from ones already seen, supports the relevant claim.

We seem to have illustrated two different types of knowledge. We can label the difference: knowledge based on a process of sampling and generalisation is *a posteriori knowledge*; knowledge based on mental arithmetic is *a priori knowledge*. More generally, our examples illustrate the truisms that a posteriori knowledge is based on sense experience², whereas a priori knowledge is in some fundamental sense independent of sense experience. On initial reflection it seems that the subjects of which we might have a priori knowledge are among the areas of central interest to philosophy: it seems that knowledge of logical and mathematical truths, moral and metaphysical principles, as well as knowledge of our own mental states, have the best chance of being shown to be a priori. The questions of whether, and how, a priori knowledge is possible are therefore of central interest to philosophy. Broadly speaking, we can gather the proposed responses into two groups, each group being associated with an influential tradition in the history of philosophy. The traditions in question differ over the emphasis they place on the senses in delivering information about the world.

Rationalists are optimistic about the possibility of knowledge about the world which is not based on the senses. Put crudely, the rationalist accepts that a priori knowledge is possible and postulates a special faculty, a capacity for "intellectual intuition", through the exercise of which we are supposed to acquire a

² By "sense experience" I mean the data delivered by the five senses as traditionally individuated. If reason was found to expand the list of the senses - perhaps by including proprioception, or because cases of telepathy were finally confirmed - then this would result in a corresponding increase in the scope of a posteriori knowledge. Knowledge through telepathy, I assume, would still be a posteriori knowledge, although nothing in this thesis depends on this assumption.

Introduction

priori knowledge. The rationalist typically supposes that the use of this faculty can yield a priori knowledge about the world: a priori knowledge of the existence of god, the relation between numbers construed as Platonic objects, and the modal structure of the universe being historically prominent examples. But the rationalist explanation of how a priori knowledge is possible has not been widely favoured in recent years³: critics of rationalism have claimed that the notion of "intellectual intuition" is mysterious, and that the postulation of such a faculty to explain a priori knowledge is *ad hoc*.

Empiricists, in contrast, reject the possibility of knowledge about the world which is not based on sense experience. Accordingly, they tend to do one of two things: accept the possibility of a priori knowledge while construing it in a way which minimises the conflict with their basic principle; or adopt a sceptical position with respect to the a priori, denying its possibility altogether, and explaining away by other means the appearance of a distinction between a priori and a posteriori knowledge.

The nonsceptical empiricist strategy for explaining how a priori knowledge is possible grows directly out of the adoption of the verificationist theory of meaning, around which early twentieth century empiricists rallied. On the verificationist theory, the meaning of a sentence is identified with the set of experiences which would indicate that it was true or that it was false. From this, a criterion for meaningfulness is drawn: if no experiences could show a sentence to be true, or false, then that sentence is meaningless.⁴ In the category of meaningless sentences, these empiricists placed imperatives, expressions of like and dislike (to which they assimilated moral propositions) and religious and metaphysical statements. The remaining meaningful sentences were divided into two groups: those on whose truth value some but not all experiences bear; and those which every experience shows to be true (those which are, in Quine's figure, "vacuously

³ Bonjour is perhaps its most vocal recent defender (Bonjour 1998).

⁴ These characterisations of the verificationist theory of meaning and criterion for meaningfulness are meant to be rough. It is extremely difficult to arrive at a formulation of these ideas which allows them to play the role that verificationists intended for them. See Ayer's 1936 comments on the formulations of the principles in the first edition of *Language Truth and Logic* for an illustration of the difficulty; Miller 1998 gives a useful survey of objections to various formulations of the principles.

Introduction

confirmed").⁵ Into the first category of meaningful sentences, these empiricists placed all and only propositions about the world, a set they took to be identical with the set of propositions which, if true, would be knowable only a posteriori. In the second category were placed all and only those truths which could be known a priori. The empiricists justified this strategy on the basis of a theory of how a priori knowledge is possible. They supposed that all a priori knowledge is at root knowledge of the meanings of words; since the meanings of words are determined by social convention, a priori knowledge is supposed to be knowledge of facts created by convention. But these facts created by convention, it was claimed, are not facts about the world: in this way, by supposing that a priori knowledge is exclusively knowledge of such conventional truths, the empiricists hoped to reconcile their acceptance of a priori knowledge with their strictures against knowledge of the world not gained through the senses. Drawing from Kant, the empiricists labelled the facts constituted by convention "analytic truths"; the empiricist project of explaining the a priori became the project of explaining how convention could make something true - that is, explaining analyticity - and then using the concept of an analytic truth to explain all cases of a priori knowledge.

This empiricist project was pretty much discredited by the 1960s, though in fact the decisive objection - in Quine's "Truth by Convention" - was already in print before the second world war. The core of the difficulties faced by the project concerned explaining how convention could create truth. But this problem was exacerbated by the failure of writers in the empiricist tradition clearly to distinguish the concepts of analyticity, necessity and apriority. Where efforts were made to disentangle these concepts, empiricist authors still held that the concept of analyticity could be used to explain both apriority and necessity. The extremely

⁵ There are also the sentences which every experience shows to be false, that is, the negations of analytic truths; it seems that these should be treated as meaningful if analytic truths are, though it is not clear whether the empiricists would have been comfortable claiming that " $2+2\neq 4$ " is meaningful. (Intuitively, they should be: this seems to be a perfectly meaningful, though necessarily false, sentence.) We should note in passing that not all verificationists agreed that the vacuously confirmed propositions should be treated as meaningful: the Wittgenstein of the *Tractatus*, from which many logical empiricists drew inspiration, treated analytic truths as meaningless, though assigned them a pragmatic value as regards the transformation of other expressions.

Introduction

strong notion of analyticity needed to play such an explanatory role was also extremely vulnerable to criticism, and after Quine's influential attacks in "Two Dogmas of Empiricism", "Carnap and Logical Truth" and *Word and Object*, found few defenders.⁶

With the collapse of the theory of a priori knowledge in terms of the concept of analyticity came the collapse of attempts to allow for a priori knowledge within empiricism. For an empiricist, the only remaining option is scepticism about a priori knowledge. For such a sceptic all knowledge is supposed a posteriori, essentially dependent on sense experience. The appearance of a distinction between a priori and a posteriori knowledge is explained psychologically: our willingness to accept that five and seven make twelve on the basis of so few observations is put down to an ingrained inability to imagine counterexamples. Our imaginations are not so limited in the case of propositions like *all humans are inquisitive*, and there we need to see a representative sample before we will accept that the possibility of error has been excluded, and that we know.

Scepticism about a priori knowledge is often closely linked with the philosophical position called *naturalism*. Some authors even define naturalism as a philosophical position which rejects a priori knowledge.⁷ But on another characterisation, naturalism is taken as the view that philosophy is not fundamentally different from science: philosophy is seen as high level, abstract science, and philosophical theories are required not to appeal to concepts which are not generally scientifically accredited. The second characterisation is obviously the more neutral with respect to the question of whether a priori knowledge is possible, and one might well wonder whether the two characterisations are really perspectives on the same position. (Certainly, if we assume that it is a necessary condition for an item of knowledge to be scientific knowledge that it be a posteriori, then the second characterisation will indeed issue in scepticism about the a priori: but this assumption does not seem obviously true.) The close association of the two

⁶ Grice and Strawson (1956) buck the trend most notably.

⁷ Devitt 1998 and Kitcher 1992, for example.

Introduction

characterisations of naturalism may be in part an historical accident. Quine is centrally responsible for the current popularity of naturalism, and he advanced his arguments in favour of naturalism on the basis of a critique of analyticity. If analyticity is not clearly distinguished from apriority, it might seem that rejection of a priori knowledge is a necessary condition of the adoption of naturalism. But to conflate apriority and analyticity would be a confusion. There might be cogent reason for supposing that naturalism and a priori knowledge are fundamentally opposed, and then again there might not be. But the question of whether it is right to think that naturalism automatically rules out a priori knowledge is an important one, and is the central concern of this thesis. On the one hand, as we have already suggested, the way in which we approach a great many philosophical questions will be influenced by whether we allow there to be a priori knowledge. On the other hand, in the current philosophical climate, many feel that a concept has to be found naturalistically acceptable if it is to be legitimately employed. To appeal to a priori knowledge without indicating how it can be naturalised seems uncomfortably close to appealing to discredited concepts such as destiny, phlogiston, astrological influence, and the like. The intention in this thesis is to support the claim that acceptance of naturalism is compatible with belief in the possibility of a priori knowledge. In my view, the arguments for thinking that naturalism must entail scepticism about the a priori are based on misconceptions of the concepts of apriority and naturalism. The purpose of this thesis is to clear away some of these misconceptions and the problems they have caused. What this thesis does not seek to do is to show that the naturalist has no difficulty in giving a theory of knowledge in general. In fact, our considerations of the nature of a priori knowledge will reveal serious problems for the project of giving a naturalistic theory of knowledge in general; the best I can do is identify these problems, not solve them on the part of the naturalist. The conclusion to be drawn is that those naturalists who believe that it is possible to give a naturalistic analysis of the concept of knowledge have no grounds for objecting to the existence of a priori knowledge.

Introduction

The argument in the thesis is structured as follows. In the last part of this chapter I draw some introductory distinctions which will structure the ensuing discussion. In **chapter two** I discuss a representative example of a theory of a priori knowledge in terms of the concept of analyticity in order to see the problems which it faces, and to highlight the way in which apriority and analyticity became entangled by the early empiricists. In **chapter three** I discuss Quine's radical empiricist critique of the concept of analyticity. I maintain that although Quine's attack is not wholly successful, it has been influential in two relevant ways. First, because it reinforces the view that a priori knowledge is unrevisable knowledge; and second, because it engenders the view that naturalism is a position which essentially rejects a priori knowledge. In **chapter four** I discuss accounts of naturalism, and pick out a formulation which seems plausible and which is not obviously inconsistent with belief in the possibility of a priori knowledge. I identify the two central reasons why naturalism has been thought to exclude a priori knowledge. The first of these stems from the conjunction of the tendency to think that a priori knowledge is unrevisable knowledge with the plausible suggestion that naturalism requires all (or, practically all) knowledge to be revisable. The second arises from doubts as to whether we can make any naturalistic sense of the idea of knowledge that is independent of experience. As a framework within which to discuss these questions, I introduce Kitcher's naturalistic theory of a priori knowledge. Kitcher approaches the concept of a priori knowledge through the concept of a priori warrant (where an a priori warrant is, roughly, a belief forming process producing a priori knowledge); he helps himself to the notion of warrant in general, and proceeds by setting conditions which a warrant must meet for it to be an a priori warrant. For Kitcher, the concept of the a priori is perfectly coherent and naturalistically acceptable - thus vindicating the view that naturalism is not automatically hostile to apriorism - but his interpretation of the idea of "experience-independence" results in a conception of a priori knowledge on which very little is known a priori. I proceed by criticising Kitcher's conception of experience-independence, and by highlighting the difficulties raised by his segregation of the theory of warrant in general from the theory of a priori warrant. In **chapter five** I

Introduction

discuss Kitcher's claim that beliefs warranted a priori are *unrevisable* - equivalently, that a priori knowledge is unrevisable knowledge. I argue that there is no reason why apriorists should be committed to the unrevisability of a priori knowledge. In **chapter six** I discuss Kitcher's claim that for a warrant to be a priori it must be *infallible*: that is, that no false beliefs are ever warranted. I argue from the point of view of general epistemology that, contrary to what many of Kitcher's critics have supposed, *all* warrant is infallible. The concept of infallibility is therefore not available for drawing a distinction between a priori and a posteriori knowledge. This, however, raises the serious question of how any naturalistic theory of knowledge can respect the requirement that all warrant be infallible. In **chapter seven** I discuss Goldman's reliabilist theory of knowledge, on which naturalistic theories of a priori knowledge are based, and argue that his latest formulation of it does not meet the requirement that warrant is infallible. In **chapter eight** I propose a modification of Goldman's framework, drawn from recent contextualist theories of knowledge, which allows him to respect the requirement of infallibility without sacrificing the main features of his theory that make it useful for formulating a naturalistic theory of a priori knowledge. I discuss whether the contextualist version of reliabilism meets the requirements of naturalism, and conclude that this is doubtful. This raises serious problems for the naturalist project in epistemology generally; but the point remains that while naturalists hold out the hope of giving a naturalistic account of knowledge they have no special reason to be hostile to a priori knowledge. In **chapter nine** I address some of the problems faced by a minimalist theory of a priori knowledge - that is, one not couched in terms of unrevisability or infallibility. I discuss and criticise two recent minimalist, naturalist, theories of a priori knowledge.

The claims I make are limited: I suggest only that the main reasons to doubt the compatibility of naturalism and apriorism stem from the assumptions that a priori knowledge must be unrevisable knowledge and that a priori warrant is distinctively infallible; once these confusions are removed it does not seem impossible that the a priori could be naturalised - provided that knowledge can be. I do not offer a worked-out theory of the a priori: my discussion primarily concerns

Introduction

what the notion of a priori knowledge ought to involve, rather than how a priori knowledge is possible. The two questions are obviously linked, but we cannot make headway with the second until we have addressed the first.

Conceptual framework

As we have already mentioned, the empiricists tended to run together the concepts of analyticity, necessity and apriority. Following Kripke, these concepts have been sharply distinguished.⁸ *Analyticity* is a semantic property of a sentence; roughly speaking, a sentence is analytic just in case its truth value is determined solely by its meaning, and not by the world. *Necessity* is a metaphysical property, foremost of propositions.⁹ A proposition is necessarily true just in case it could not possibly be false. *Apriority* is an epistemological concept, concerning the conditions under which knowledge arises. But it is not immediately obvious just of what apriority is a property; for there are two ways of construing the following sentence:

S knows *p* a priori.

On one reading, apriority is taken to be a property of a proposition: on this view, the sentence above should be read as claiming:

S knows *p*, and *p* is an a priori proposition.

On the other reading, apriority is taken to be a property of a way of knowing: on this view, the sentence "S knows *p* a priori" should be read as claiming:

S knows *p* in an a priori way.

If we adopt the first interpretation of "S knows *p* a priori", an explanation of apriority will be an explanation of what property a proposition must have for it to be a priori. If we adopt the second reading, our goal will be to determine which

⁸ Kripke 1980.

⁹ I pass over discussion of essential properties.

Introduction

ways of knowing are a priori ways of knowing. The way in which we conduct our investigation into a priori knowledge is therefore highly sensitive to our decision as to what "a priori" attaches to. It is arguable, however, that the second conception must be fundamental, that is, that apriority is fundamentally a property of a way of knowing. For, let $p \vee q$ be a compound proposition consisting of the disjunction of a mathematical truth p , and an empirical truth q .¹⁰ Knowledge of either p or q allows us to infer $p \vee q$ by disjunction introduction, and it is very plausible that a logical operation such as disjunction introduction preserves the epistemic status of the input beliefs: that is, if p is known a priori, then $p \vee q$ will be known a priori; if p is known a posteriori then $p \vee q$ will be known a posteriori. So the epistemic status of the disjunction will depend on which of the disjuncts it is derived from: we have a single proposition that can be known either a priori or a posteriori, depending which inferences the belief in it is based on. But if the status of the disjunctive proposition can vary without any change in the proposition itself, what determines its epistemic status cannot be a feature of the proposition.

This, then, is strong evidence that "a priori" is fundamentally a property of a way of knowing, not of a proposition. To say "S knows p a priori" is to say that S knows p in an a priori way: to prefigure terminology that we have not yet properly introduced, it is to say that S's belief in p is a priori warranted. If we had an account of what it is for a way of knowing to be a priori we would be able to define a derivative notion of what it was for a proposition to be a priori: however, this is a matter of some delicacy. As Kripke pointed out, it will not do to say that a proposition is a priori just in case it can *only* be known in an a priori way. For we must bear in mind that to claim that an item of knowledge is a priori does not rule out the possibility that it could be known through methods more usually associated with a posteriori knowledge. If we are lazy, we can use a calculator to solve arithmetic problems which we could have done in our heads with a little effort. As Kripke pointed out, the arithmetical knowledge derived from the calculator is best regarded as a posteriori knowledge, for our knowledge that the calculator has produced the right answer is based on our knowledge of the way the machine was

¹⁰ For example: $(2+2=4) \vee (\text{grass is green})$. This argument is based on Kripke 1981: 35.

Introduction

constructed, the physical laws on which it depends, etc. But this sort of knowledge is a posteriori knowledge, and it seems right to say that knowledge based on a posteriori knowledge will also be a posteriori.¹¹

The right way to define a derivative notion of the apriority of a proposition is to say that a proposition is a priori just in case it *could* have been known a priori. That is, the idea of apriority construed as a property of a way of knowing is related to the idea of the apriority of a proposition by the following equivalence:

$$p \text{ is a priori iff } \Diamond \exists S(S \text{ knows } p \text{ a priori}).$$

This introduces the problem of specifying just what is meant by "possibly" in this context. What is impossible for a human may be possible for a Martian, but it is not at all clear whether we would want to say that a proposition was a priori if it could be known a priori by a Martian but not by us. Who knows what Martians can know a priori? These are difficulties which we will put aside, however. For the purposes of this thesis we will focus on the idea of an a priori way of knowing, reverting to talk of a priori propositions only when to do otherwise would cause too many problems in our discussion of an author.

This completes the introduction of distinctions which the argument of this thesis will take for granted. We have seen that the concepts of apriority, necessity and analyticity are distinct¹², and that the notion of a way of knowing being a priori is more fundamental than the idea of a proposition being a priori. In the next

¹¹ Kripke 1981: 35.

¹² I have passed over discussion of proposed instances of necessary a posteriori, and contingent a priori, truths. Although I am strongly inclined to think that there are such propositions, the point that necessity and apriority are distinct *concepts* does not hang on whether proposed examples are upheld. The property of being necessary is distinct from the property of being knowable a priori, just as the property of being a creature with a heart is distinct from the property of being a creature with kidneys.

Introduction

chapter we will look more closely at the empiricists' attempt to explain a priori knowledge, and the difficulties associated with clarifying the notion of analyticity.

CHAPTER TWO

Carnap's account of a priori knowledge

Chapters two and three are concerned to show that Quine and Carnap, despite their differences regarding the viability of the concept of analyticity, agree on one crucial issue: that an appeal to this concept is essential to the explanation of a priori knowledge. (We will call the project to provide an explanation of a priori knowledge in terms of analyticity, the project to find an *analytic theory* of a priori knowledge.) In the chapters that follow, much work will be done to rid the concept of a priori knowledge of the overtones it acquired through liaison with the analytic theory; but the present purpose is to describe that theory, and the controversy over its central concept: analyticity. We begin with an overview of empiricist theories of a priori knowledge.

There are three distinct empiricist strategies for dealing with the problem of a priori knowledge.¹ The first of these, associated in particular with British Empiricists such as Locke and Hume, need not detain us, being in effect no strategy at all. This account affirms the importance of experience for knowledge, while simultaneously recognising a class of a priori knowable propositions² (most famously recorded in Hume's distinction between "matters of fact" and "relations of ideas"). However, no serious attempt is made to provide an epistemology for these propositions: they are dismissed as "trivial" and the problem of accounting for our knowledge of them in a way consistent with empiricism ignored.³ There is, therefore, little to interest us here as regards theories of a priori knowledge.

The second strategy is associated primarily with the logical empiricism which arose out of the Vienna Circle. Accounts of a priori knowledge which adopt this strategy have been termed "moderate empiricist".⁴ The strategy here is to account for a priori knowledge by co-opting Kant's notion of analyticity. In the

¹ I ignore Mill's thoroughgoing empiricism (or "low inductivism", in Kim's phrase) about logical and mathematical knowledge. Mill's low inductivism is not generally considered viable (see Kim 1981).

² A general definition of the apriority of a proposition is given in chapter one.

³ Dancy 1985.

⁴ The terminology is drawn from Bonjour 1998.

Carnap's account of a priori knowledge

standard formulation, a sentence is analytic just in case it is true solely in virtue of its meaning and independently of matters of fact. Logic and mathematics being the most prominent bodies of a priori knowledge, the moderate empiricist hopes in particular to show that the propositions thereof are analytic; more generally, it is hoped that the concept of analyticity can be explained in such a way as to substantiate the claim that all and only a priori propositions are analytic propositions.⁵

The third strategy characterises the radical empiricist account developed by Quine; this account arises out of Quine's famous critique of moderate empiricism, advanced particularly in his "Two Dogmas of Empiricism", "Carnap and Logical Truth", and "Truth by Convention". Quine develops an epistemology in which the existence of a priori knowledge is specifically excluded.

The discussion in this chapter concerns accounts of the second type; chapter three discusses Quine's arguments for radical empiricism. The intention is not to give an exhaustive catalogue of different notions of analyticity and their failings; rather, I am primarily concerned to trace the dialectic running from the adoption of the analytic theory to scepticism about a priori knowledge.

Moderate empiricism: the analytic theory of a priori knowledge

For the purposes of dramatising the challenge posed to empiricism by the existence of a priori knowledge, we can gloss the central principle of empiricism in the following slogan: *all knowledge is dependent on experience*. But a priori knowledge is experience-independent knowledge. Hence it is crucial for those empiricists who wish to retain the intuition that certain propositions are knowable a priori - as do the moderate empiricists - that they find a way to reconcile the existence of such propositions with their general principles. Dancy suggests that the general idea behind moderate empiricist accounts is to relax the requirement that all knowledge is dependent on experience, requiring instead that all *factual* knowledge is

⁵ Cf. BonJour 1992.

Carnap's account of a priori knowledge

dependent on experience.⁶ A priori knowledge, the suggestion goes, is not to be thought of as knowledge of facts.

On the most literal reading of this suggestion, it appears as an injunction to treat putative a priori propositions as non-truth-evaluable. Such an account can be called *noncognitivist*. An approach like this may be familiar from moral philosophy: some moral philosophers have held that putative moral judgements are really no more than disguised expressions of like or dislike. However plausible this strategy may be in moral philosophy⁷, it is not an option that will be considered here. A noncognitivist reading of a priori knowledge would run strongly counter to our intuitions that there can be genuine truth and falsehood in logic and mathematics. Noncognitivist accounts of logical and mathematics have been offered in the past: Wittgenstein in the *Tractatus* denies that the theorems of propositional logic have meaning, and asserts that mathematical truths should be read as rules for the transformation of expressions. However, such a reading should be considered a variety of scepticism about a priori knowledge. Knowledge is by definition always of truths; *a fortiori* a priori knowledge is always of truths. Where there is no truth, there can be no knowledge and therefore no a priori knowledge. The noncognitivist interpretation of Dancy's suggestion should therefore be rejected.

Dancy's suggestion should rather be interpreted as a call for a division within the set of all facts. The moderate empiricists' proposal is that this distinction is to be drawn on semantic grounds. This is where the concept of analyticity is meant to help. The empiricists suppose that it is intuitively obvious that there are analytic sentences, that is, sentences which are true just in virtue of what they mean and independently of matters of fact. Quine suggests that the intuition that the inclination to believe in analytic truths⁸ arises naturally out of the truism that the truth of a sentence depends jointly on meaning of the sentence (that is, the proposition it expresses) and the facts.⁹ If the word "killed" meant begat, Quine points out, then the sentence "Brutus killed Caesar" would be just as false as if, reverting to normal meanings, Caesar had not been killed by Brutus. Given this

⁶ Dancy 1985.

⁷ It is not very plausible: it faces the "Frege-Geach" problem (Cf. discussion in Miller 1998).

⁸ An "analytic truth" is a true sentence which is analytic.

⁹ Quine 1953: 36; 1966: 101; 1970: 95.

Carnap's account of a priori knowledge

double dependence, there is a natural inclination to think that there could be expressions in which the factual component is *nil*: and these are the analytic truths. The idea of an analytic proposition is meant to be an intuitive one, and it seems that that appeal to this concept has the potential to avoid the problems with noncognitivism: analytic truths are indeed true, but their truth is thought to be fixed in some special way.

Analyticity is, therefore, a semantic concept which is pressed by the moderate empiricists into service in an epistemological role. A successful analytic theory of a priori knowledge would have two features. First, it would entail the truth of the following equivalence:

$$(1) \quad \forall p (p \text{ is a priori} \leftrightarrow p \text{ is analytic}) .^{10}$$

Second, it would provide a satisfying explanation of why (1) was true. A key test for the analytic theory will be whether the truths of logic and mathematics can be shown to be analytic.

It is worth noting, albeit in passing, that the moderate empiricists often appealed to the concept of analyticity in a second capacity: that of providing an *analytic theory of necessity*. To a certain extent, no doubt, this strategy was adopted through failure clearly to distinguish the concepts of necessity and apriority. There may, though, have been a more legitimate motivation for the theory. For to many modern philosophers it has seemed that knowledge only of the material truth of a proposition was given in experience; knowledge of the modal status of a proposition was not thought to be so given. So it seems that knowledge that a proposition is necessarily true must be a priori knowledge. This might seem to give reason to adopt the following claim¹¹:

¹⁰ Analyticity is a semantic property and therefore attaches to sentences, not propositions. The apparent predication of analyticity of a proposition in (1) is shorthand for the claim that the proposition is expressed by a sentence which is analytic.

¹¹ Not good reason, though: the biconditional (2) claims that all and only necessary truths are knowable a priori; whereas the rough considerations offered in support of (2) were to the effect that knowledge *that* a proposition was necessarily true had to be a priori. The biconditional talks of the

Carnap's account of a priori knowledge

- (2) $\forall p (p \text{ is necessary} \leftrightarrow p \text{ is a priori})$.

The moderate empiricists' assumption of the analytic theory of the a priori then lends support to the idea of an analytic theory of necessity, since (1) and (2) together give us

- (3) $\forall p (p \text{ is necessary} \leftrightarrow p \text{ is analytic})$.

For philosophers sceptical about necessity (such as Quine) this provides an opportunity to bring anti-metaphysical arguments to bear against the view that there are analytic truths. However, as noted in the introduction, the chief concern in this thesis is epistemological and from here on the analytic theory of necessity will be set on one side.

The account of analyticity offered so far is vague and unenlightening: we have no idea how analyticity so defined is meant to help explain the possibility of a priori knowledge. To discuss exhaustively concepts of analyticity would be a large and largely fruitless task: a great many versions of the concept of analyticity appear in the literature, and the distinctions between them are not always recognised by the authors themselves (Ayer's chapter "The A Priori" in *Language, Truth and Logic* is an example of a particularly bad case of conflation of non-equivalent definitions).¹² In assessing the moderate empiricist programme I propose to concentrate on Carnap: first, because his account of analyticity is well worked-out and its shortcomings therefore clearer; and second, because the chief critic of analyticity, Quine, focuses on Carnap's writing. The goal of the discussion of Carnap is to find an account of analyticity which will vindicate the analytic theory of a priori knowledge. The search will end in failure, but the process of seeing it fail will highlight the conceptual independence of apriority and analyticity, and will place

knowledge of a proposition; the support for the biconditional talks of the knowledge of the *modal status* of a proposition. But there is no need for us to address this problem here.

¹² The task of cataloguing and criticising a wide variety of definitions of analyticity is well carried out by BonJour 1993/1998. BonJour, however, shows little sympathy for empiricism.

Carnap's account of a priori knowledge

us in a position to observe Quine's own tacit denial of this difference in his own assumption of the analytic theory.

Carnap's account of language and analyticity

Carnap tries many times to give an account of analyticity which will vindicate the analytic theory. Briefly stated, his most worked out attempt to explain the concept of analyticity appeals to what he sees as the conventional nature of syntactical linguistic rules (that is, rules for the formation of linguistic expressions): a sentence will be analytic just in case it is provable purely on the basis of such rules. In this way, Carnap hopes to give an epistemology for logic and mathematics which does not appeal to obscure rationalist principles such as those adopted by Russell¹³, and which we discussed briefly in chapter one. Conventions are chosen on pragmatic, not evidential, grounds: this, it is hoped, will give conventions the independence from experience needed to explain how all and only a priori truths are analytic. In what follows, Carnap's proposals will be looked at in more detail, and we will see that the deficiencies of the account of analyticity in terms of conventional syntactic rules drives him to introduce a new, less perspicuous notion of analyticity. In trying to show that this latter notion supports the analytic theory of a priori knowledge, he has to assume that logical and mathematical truths are analytic in this revised sense. This assumption will be seen to vitiate his attempt to establish the analytic theory.

Carnap thinks the notion of analyticity is intuitively clear. He believes that his theory of language reconstructs the concept and successfully assigns it a technical role. In common with many authors of his period Carnap holds that natural language is hopelessly vague and riddled with confusion: for scientific work, he asserts, it is necessary to construct new languages which meet adequate standards of clarity.¹⁴ To speak of a class of entity *x*, we need to set up *meaning rules*

¹³ This point is made by Creath 1992.

¹⁴ Most notably, perhaps, Tarski (Tarski 1956).

Carnap's account of a priori knowledge

for forming and transforming expressions containing references to x . These rules compose *linguistic frameworks*; such frameworks can deal with any kind of entity or phenomenon we choose, whether abstract or physical. The apparatus of such frameworks allows Carnap to identify what he takes to be an ambiguity with respect to questions of existence. The question "does x exist?" can be interpreted in two ways: as an *internal* or as an *external* question. Interpreted as an internal question, it is resolvable using the meaning rules of the language. Questions about the nature of the things which exist can likewise be decided by appeal to the rules, or, where the language deals with physical phenomena and the question concerns empirical facts, through a combination of application of the rules and empirical investigation. Interpreted as an external question, however, the situation is different; read as an external question Carnap holds that the question "does x exist?" can only be understood as "is it reasonable to adopt a linguistic framework in which there are x s?"¹⁵ This is crucial for Carnap's analytic theory of a priori knowledge: for the central tenet of Carnap's theory of a priori knowledge is that questions regarding the reasonableness of adopting a given linguistic framework are pragmatic questions. It is reasonable to adopt a framework which contains a given group of entities just in case the adoption of that framework will efficiently and fruitfully perform the tasks for which it was developed: such a task might be the communication of knowledge¹⁶, or, for a scientific language, the prediction and explanation of phenomena. Carnap's distinction between internal and external questions allows him to classify metaphysical questions as nonsense: questions such as "are there physical objects?" or "are there numbers?" are paradigm cases for disambiguation as just outlined. Taken as internal questions, there are physical objects or numbers just in case the language we adopt contains the apparatus for constructing expressions which contain symbols for such entities. In Carnap's opinion, however, such questions are not normally meant as internal ones, nor are they meant as pragmatic questions concerning the adoption of linguistic frameworks; in which case, he writes, the questions do not have a clear cognitive

¹⁵ Carnap 1956: Appendix A *passim*, though esp. 206.

¹⁶ *Ibid.*: 208.

Carnap's account of a priori knowledge

content: they are "pseudo questions".¹⁷ However, while maintaining this strict distinction between internal and external questions, Carnap suggests that we should be generous in allowing the development of linguistic frameworks. It is a philosopher's job to assess frameworks for consistency, carefully to distinguish internal from external questions, and to resolve those internal questions which can be resolved purely through application of the meaning rules, but where that application is complex. A philosopher has no say over which linguistic frameworks should be adopted, since this is a matter for decision by those who will use them. Carnap expresses this in his *Principle of Tolerance*.

*It is not our business to set up prohibitions, but to arrive at conventions. [...] In logic, there are no morals. Everyone is at liberty to build up his own logic, i.e. his own form of language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments.*¹⁸

The rules of a linguistic framework are to be considered *implicit definitions* of the expressions that they introduce. A sentence A of some language L is an implicit definition just in case A cannot be false without becoming a sentence of a language different from L ; equivalently, the truth of A is constitutive of the meaning of a term that it involves.¹⁹ We implicitly define a term by assigning truth values to sentences involving the term. If the implicit definition is successful, the term in question will thereby be assigned whatever meaning is necessary for the sentences in which it occurs to be true. As a consequence, linguistic frameworks cannot directly be compared. Where some name " t " is defined one way in a language L_1 , and defined differently in another language L_2 , L_1 and L_2 do not conflict over the nature of ts , because " t " in L_1 refers to something different from " t " in L_2 . To adapt an example of Carnap's, we can consider two languages, one of which defines the spatial metric g_{uv} as a constant, the other as a variable.²⁰ These two languages do not disagree over the nature of the spatial metric; rather, the expression " g_{uv} " means something

¹⁷ *Ibid.*: 209. See also Carnap 1937: 281ff.

¹⁸ Carnap 1937: 51-2.

¹⁹ Cf.: Boghossian 1997: 348.

²⁰ Carnap 1937: 178-9.

Carnap's account of a priori knowledge

different in each language. So the claim " g_{uv} is a constant" can be made truly in the first language, and the claim " g_{uv} is a variable" made truly in the second language, without contradiction.

However, there are two less direct senses in which linguistic frameworks can be compared.²¹ First, they can be compared with respect to their practical utility in discharging the tasks for which they were constructed. Secondly, they can be compared in point of how well the concepts they contain match intuitive concepts. For Carnap, where a constructed concept has a meaning sufficiently close to that of an intuitive concept already in use, the constructed concept can be considered an *explication* of that intuitive concept. Although failure by a framework to provide explication of intuitive concepts would not tell against that framework, frameworks can be compared indirectly with respect to the degree to which they provide explications of concepts already in use.

As suggested earlier, Carnap holds that his system of linguistic frameworks allows him to explicate the intuitive concept of analyticity. His most developed attempt to provide such an explication is that presented in *The Logical Syntax of Language*.²² Ideally, Carnap would like a sentence to be analytic just in case it is provable solely on the basis the syntactic rules of the linguistic framework.²³ The syntactic rules are supposed to be conventions for communication, adopted on pragmatic, not evidential, grounds. Knowledge that a syntactic rule holds is therefore supposed to be knowledge based not on experience, but on grasp of what conventions have been adopted; and this independence from experience is meant to transfer to all the theorems of the rules of the framework. If a sentence is provable by reference to the syntactic rules, knowledge of it is at root knowledge of conventions, and hence independent from experience, and hence a priori.

For Carnap's conventionalist account to be distinguished from nonfactualism, it is important we establish that analytic sentences on Carnap's

²¹ Creath 1992: 145.

²² Carnap 1937.

²³ Carnap 1937: 39ff, 182ff.

Carnap's account of a priori knowledge

attempted model do in fact express genuine propositions, capable of being true or false; if they are not truth-evaluable, a priori propositions would not be truth-evaluable either, and we have argued that this position amounts to scepticism about the existence of a priori knowledge.²⁴ Clearly, for some analytic sentence p in a linguistic framework L , Carnap would treat the question "is p true?" as a paradigm case for disambiguation via the distinction between internal and external questions. Treated as an internal question, the answer to the question is clearly "yes"; treated as an external question, the question must be taken to mean "are there pragmatic reasons for adopting L ?", to which specifically philosophical considerations are not relevant. Therefore, within Carnap's own framework, analytic sentences clearly possess truth values.

But it is not necessary to lean too heavily on Carnap's own system to give a positive answer to the question of whether analytic sentences regarded as expressions of conventions are in fact true. As noted above, the rules of the linguistic frameworks are to be considered implicit definitions of the terms they introduce; that is, the terms gain their meaning through conventional decisions regarding the truth values of the sentences in which they occur. This procedure therefore involves assigning truth values to sentences: but the bearers of truth values are propositions, and a sentence cannot have a truth value unless it expresses a proposition. The procedure of implicit definition, therefore, guarantees that the admissible sentences of a linguistic framework will express propositions.

Carnap's account therefore seems able to explain our a priori knowledge of logical and mathematical propositions through appeal to a concept of analyticity cashed out in terms of conventions, and in a way that avoids noncognitivism. However, his account faces two problems.

First, it is incomplete. In a linguistic framework with the resources to express facts about the physical world, there will be syntactical rules which are designed for the formation of expressions which state empirical truths. If provability by a linguistic rule was all there was to Carnap's account, these rules

²⁴ See page 15.

Carnap's account of a priori knowledge

and the sentences derivable from them would have to be considered analytic truths. But, of course, one of Carnap's goals in setting up his system of linguistic frameworks is to explain the intuitively privileged epistemic status of truths of mathematics and logic²⁵; his goal will not be achieved should his version of the analytic theory have the consequence that intuitively empirical facts have the status of analytic truths. He therefore needs to identify, within the set of sentences provable on the basis of linguistic rules, a distinction which corresponds to the intuitive distinction between logico-mathematical truths and empirical truths.

An example might make the problem clearer. Consider a linguistic framework set up for the purposes of expressing facts about Newtonian dynamics. Such a linguistic framework, we might suppose, would be composed of rules allowing the formation of expressions which, suitably interpreted, would express logical and mathematical truths, and also facts involving properties of spatiotemporal location, mass, velocity and the relationships set up between these properties by the Newtonian framework. Some of these rules, let us suppose, allow symbols to be combined to create the formula: "force=mass \times acceleration". Other rules will allow the combination of symbols to create the formula "2+2=4". From a syntactic point of view, both of these formulae will be on a par: both are derivable from the rules of the framework. However, given the obvious interpretation, the latter intuitively will be, while the former intuitively will not be, an a priori truth. If Carnap's system of frameworks is to explain our intuitions about the special epistemic status of mathematical propositions he needs to identify some principled difference between these two.

The second problem is particularly pressing: Carnap's system of linguistic frameworks is unable in principle to account for the epistemic status of logic itself. This important result - that logic cannot be true by convention - was established by Quine in his 1936 paper "Truth by Convention".²⁶ The significance of Quine's result warrants an extended discussion of his paper; afterwards, we will see how Carnap

²⁵ Stressed by Friedman, in Aspray and Kitcher 1988.

²⁶ Reprinted in Quine 1966: 70-99.

Carnap's account of a priori knowledge

modified his system of linguistic frameworks so as to limit the damage to his project which arises from Quine's argument, and how he tried to address the first problem.

Quine's "Truth by Convention"

Quine's 1936 paper is concerned to distinguish two types of convention. The first of these can be called *notational convention*, and is the subject of section I of the paper. Here, Quine discusses the possibility of constructing mathematics out of logic through defining mathematical terms in terms of a small set of logical terms; to be adequate, the definitions must proceed in such a way that the prior truth values of mathematical sentences remain unchanged by the transformation into sentences containing only logical vocabulary. The project of demonstrating that mathematics is so constructible out of logic through definitions is known as the *logician* programme. Quine opines that the results of Russell and Whitehead's *Principia Mathematica* demonstrate that the logicist programme poses no technical problems, although the project was yet to be completed; in fact, there turned out to be significant obstacles in the way of the completion of the logicist programme but these do not here concern us. Of most interest for present purposes is the status of the definitions which reduce mathematical expressions to expressions containing only logical vocabulary; these definitions are conventions of notation. Quine writes:

If for the moment we grant that all mathematics is thus definitionally constructible from logic, then mathematics becomes true by convention in a relative sense: mathematical truths become conventional transcriptions of logical truths.²⁷

The element of convention enters here since, assuming that all mathematics is expressible in a purely logical vocabulary, we avail ourselves of non-reduced mathematical expressions purely for the purpose of convenience: simple numerical expressions are easier to work with than their complex logical articulations. The

²⁷ Quine 1966: 80.

Carnap's account of a priori knowledge

definitions supporting the construction of mathematics from logic are therefore conventions for *transforming* truths; what Carnap requires, however, are conventions for *creating* truths.²⁸ Conventions of notation can confer on mathematics the epistemic status of logic; however, such conventions cannot help to explain the epistemic status of logic itself. Trivially, logic is reducible to logic through conventions of notation: no epistemic insight to the nature of logic is afforded by this observation. If appeal to conventions is to afford insight into the epistemic status of logic, then, a different sort of convention will have to be employed; we can call this *truth by convention* proper. Section II of Quine's paper discusses the possibility that logic can be thought of as conventional in this sense.

Section II considers the problem of how a reduced set of logical axioms can be rendered true by convention. Quine selects for discussion axioms governing the operators for negation, material conditional and universal quantification. It can be shown that these operators are adequate for generating a complete system of quantified first-order logic²⁹; if, through the conventional introduction of linguistic rules which assign them meaning, these concepts can be reconstructed within a linguistic framework in such a way as to capture their customary usage, it may seem reasonable to assert that the theorems of quantified first order logic are true by convention. As we have seen in our earlier discussion of implicit definition, given an undefined symbol "*t*" we can implicitly define it through assigning truth values to certain expressions containing "*t*". Given uninterpreted symbols "not", "if...then..." and "for every...", we are to assign them a meaning by specifying which formulae containing them are true. This will be done by assigning truth values in the first instance to formulae containing only the logical symbols essentially. However, as Quine notes, we cannot simply list the expressions which are to be true.

It would appear that we sit down to a list of expressions and check off as arbitrarily true all those which, under ordinary usage, are true statements involving only our logical primitives essentially; but this picture wanes when we reflect that the number of such statements is infinite. If the convention whereby

²⁸ *Ibid.*: 81.

²⁹ See for example Hunter 1996.

Carnap's account of a priori knowledge

those statements are singled out as true is to be formulated in finite terms, we must avail ourselves of conditions finite in length which determine infinite classes of expressions.³⁰

This requirement will be met by appealing to axiom schemas. The intention is to appeal to axioms already in use, and to exploit these in order to assign truth to infinite classes of expressions containing the as-yet undefined logical symbols essentially. Quine takes as an example the schema

- (i) If if p then q then if if q then r then if p then r .³¹

An infinite class of expressions are then assigned the value *true* by the adoption of the following convention:

- (I) Let all results of putting a statement for ' p ', a statement for ' q ', and a statement for ' r ' in (i) be true.³²

Two further conventions suffice to fix conventionally the meaning of the symbols "not" and "if....then..." in such a way as to reflect ordinary usage, and generate a system of propositional logic. These are (taking the symbol " \sim " to stand for the symbol "not"):

- (II) Let any expression be true which yields a truth when put for ' q ' in the result of putting a truth for ' p ' in 'If p then q '.
- (III) Let all results of putting a statement for ' p ' and a statement for ' q ', in 'If p then if $\sim p$ then q ' or 'If if $\sim p$ then p then p ', be true.

Quine observes that the three conventions generate all and only expressions which under ordinary usage are truths involving only the logical operators "if...then..." and "not" essentially.³³ All of these expressions are thus rendered true by convention;

³⁰ Quine 1966: 84-5.

³¹ This corresponds to the principle of transitivity: $p \rightarrow q, q \rightarrow r \vdash p \rightarrow r$.

³² *Ibid.*: 85. (Note: Quine refers to our proposition (i) as (1)).

³³ *Ibid.*: 89. The proof of this is a variant on the completeness proof for the propositional calculus. For details see Hunter 1996.

Carnap's account of a priori knowledge

and the derivation of theorems can be considered the unfolding of the conventions we have adopted.

The system of propositional logic thus generated could be expanded into a system of quantified logic. This would involve the adoption of further conventions governing the assignment of truth to expressions containing the symbol "every", and so implicitly defining that symbol, but we need not enter into the details of this here. However, as Quine points out in section III, the expansion need not stop there. Geometry can be rendered true by convention, by conventionally assigning truth to the axioms of a suitably axiomatised system of geometry.³⁴ And further conventions could be set up which assigned to symbols such as "mass", "velocity" etc., a meaning which matched the meaning already in use for these symbols, rendering empirical propositions true by convention as well. Quine concludes:

If in describing logic and mathematics as true by convention what is meant is that the primitives *can* be conventionally circumscribed in such fashion as to generate all and only the accepted truths of logic and mathematics, the characterisation is empty; our last considerations show that the same might be said of any other body of doctrine as well.³⁵

This of course is just the problem with Carnap's system of linguistic frameworks which we mentioned not far back: in order to give substance to the claim that the special epistemic status of logic and mathematics is to be accounted for by appeal to conventions, he has to identify a principled difference between logico-mathematical truths construed conventionally and empirical truths so construed. Carnap's attempt to locate such a difference will be turned to shortly; Quine's conclusion in the final pages of his 1936 paper must be set out first, as it seriously affects the strategies available to Carnap in drawing the required distinction.

In the last five pages of the 1936 paper, Quine addresses the possibility that logic can be considered true by convention in the sense that conventions actually create the truths of logic. As we noted, there are an infinity of contexts in which the logical connectives can appear, and therefore the conventions adopted in assigning

³⁴ Quine appeals to Huntington's axiomatisation of geometry, (Huntingdon, E.V, 1913. "A Set of Postulates for Abstract Geometry", *Mathematische Annalen* 73, cited in Quine 1966: 77)

³⁵ Quine 1966: 95.

Carnap's account of a priori knowledge

meaning to the logical constants must assign truth to infinite classes of expressions at once. The conventions will therefore be general. But the application of a general convention to a specific case will require an inference; and this inference must appeal to the logical concepts implicitly defined by the conventions themselves.

To make this problem clear, we shall follow through Quine's example of inference construed as the unfolding of conventions.³⁶ Quine's example is as follows. From

(3) If time is money then if time is not money then time is money

and

(5) If if time is money then if time is not money then time is money then if if time is not money then time is money then time is money then if time is money then time is money

we can derive:

(6) If if if time is not money then time is money then time is money then if time is money then time is money .³⁷

The convention expressed in (II) requires that we adopt as true any expression which yields a truth when put for '*q*' in the result of putting a truth for '*p*' in 'If *p* then *q*'. (6) has this property. Therefore, (6) is true by convention also. That is, (II), in conjunction with (3) and (5), directs us to adopt the convention:

(7) (6) is to be true.

³⁶ I have renumbered the propositions in Quine's example to be continuous with the numbering of propositions in this chapter overall.

³⁷ Quine's example is hard to follow; but by writing "P" for the sentence "time is money", and inserting brackets where appropriate, the structure of the inference becomes plain:

(3) $P \rightarrow (\sim P \rightarrow P)$

(5) $(P \rightarrow (\sim P \rightarrow P)) \rightarrow (((\sim P \rightarrow P) \rightarrow P) \rightarrow (P \rightarrow P))$

(6) $((\sim P \rightarrow P) \rightarrow P) \rightarrow (P \rightarrow P)$.

Carnap's account of a priori knowledge

This inference, if construed as the unfolding of conventions, masks an infinite regress. For the inference of (6) from (3) and (5) must be interpreted as an inference from the general convention (II), plus the premise that (3) and (5) are true, to (7). To expose the regress, Quine proposes an uncontentious rewriting of (II) as follows:

- (II') *No matter what x may be, no matter what y may be, no matter what z may be, if x and z are true [statements] and z is the result of putting x for 'p' and y for 'q' in 'If p then q' then y is to be true.*

The additional premise, required for the deduction of (7), is:

- (8) (3) and (5) are true and (5) is the result of putting (3) for 'p' and (6) for 'q' in 'If p then q'.

From (II') and (8) we can infer (7). Quine:

This inference is obviously sound logic; as logic, however, it involves use of (II') and others of the conventions from which logic is supposed to spring.³⁸

Following through this inference on the basis of the conventional account of logic reveals the regress. Suppose we have a convention (IV), introduced as an implicit definition of the symbol "every", which allows us to perform universal instantiation - that is, to infer specific instances from universally quantified expressions. Three applications of convention (IV) on (II') allow us to infer the following:

- (9) If (3) and (5) are true and (5) is the result of putting (3) for 'p' and (6) for 'q' in 'If p then q' then (6) is to be true.

Then (7) can be inferred from (8) and (9), via the following application of convention (II'):

- (10) (8) and (9) are true and (9) is the result of putting (8) for 'p' and (7) for 'q' in 'If p then q'.

³⁸ *Ibid.*: 96.

Carnap's account of a priori knowledge

But to infer (7) from (10) and (II') we have to go through just the procedure needed to get (6) from (8) and (II'). The regress is evident. Quine concludes:

In a word, the difficulty is that if logic is to proceed *mediately* from conventions, logic is needed for inferring logic from the conventions.³⁹

The difficulty can also be seen by focusing on the requirement, mentioned briefly earlier, that the symbols to be conventionally defined as logical operators start off uninterpreted; to interpret the symbols we need recourse to a metalanguage which already contains logical apparatus. Therefore, first order predicate logic cannot be true by convention.⁴⁰

Carnap's response

This result poses serious problems for the moderate empiricist attempt to motivate (1). Intuitively, logical truths are knowable a priori; the demonstration that knowledge of logical truths cannot be explained by appeal to convention thus puts an infinite class of propositions outside the remit of the conventionalist programme. The force of Quine's argument is such that moderate empiricist responses to it are essentially exercises in damage-limitation. Often this damage limitation amounts to little more than a retreat into obfuscation: it is common, for example, in introductory books on philosophy to find logical truths described as "true in virtue of what they mean" without any effort to explain how this appeal to analyticity is supposed to help. This is not the approach taken by Carnap, however, who in *The Logical Syntax of Language* assigns quantified first order logic a special status. Carnap's strategy here has been well set out by Friedman.⁴¹ The idea was to

³⁹ *Ibid.*: 97.

⁴⁰ In asserting this conclusion I pass over the possibility of giving a conventional account of logic whereby logic proceeds *immediately* by convention. In an influential article, Dummett interprets Wittgenstein as offering such an account (Dummett 1959). However, this type of *radical conventionalism* has not been thought to be promising, and I will not discuss it. (See also Putnam 1983).

⁴¹ Friedman in Aspray and Kitcher 1988.

Carnap's account of a priori knowledge

respond to Quine's demonstration that no conventionalist epistemology could be given for quantified first order logic by showing that no such epistemology was needed. Friedman writes that Wittgenstein, in the *Tractatus*, attempted to show that the logic of Frege's *Begriffsschrift* (that is: first order quantified logic) is built into thought itself, being built into any system of representation that could be called a language.

...from Wittgenstein's point of view, the *Begriffsschrift* rests on two basic ideas [sic]: Frege's function/argument analysis of predication and quantification, and the iterative construction of complex expressions from simpler expressions via truth-functions. So any language in which we can discern both function/argument structure - in essence, where there are grammatical categories of intersubstitutable terms - and truth-functional iterative constructions will automatically contain all the logical forms and principles of the new logic as well. Since it is plausible to suppose that any system of representation lacking these two features cannot count as a language in any interesting sense, it makes perfectly good sense to view the new logic as delimiting the general conditions of any rational thinking whatsoever. For the new logic is now seen as embodying the most general conditions of meaningfulness (meaningful representation) as such.⁴²

On this view, Friedman writes, it makes no sense to sense to enquire into the foundations of logic:

Logic rests on no facts whatsoever, and certainly not facts about the meaning or usages of English (or German) words. Rather, logic rests on the abstract combinatorial possibilities common to all languages as such. In this sense, logic is absolutely presuppositionless and thus absolutely uncontentious.⁴³

Given that Carnap, as a moderate empiricist, is committed to producing an analytic theory of a priori knowledge, we should read this strategy as an attempt to shore up the theory that logical truths are analytic in the face of Quine's critique of conventionalism. But the sense in which logic thereby becomes "analytic" is, in spite of Friedman's comments, very unclear. The analyticity of logical truths, it seems, is supposed to stem from logic being built into any possible language. It is

⁴² *Ibid.*: 84.

⁴³ *Ibid.*: 91.

Carnap's account of a priori knowledge

perhaps intuitively plausible to say that such a view attributes analyticity to logical truths; certainly they are "true in virtue of meaning" in some sense. But the sense in which they are true *in virtue of* meaning is hard to precisify and evaluate. We shall shortly see, in Quine's "Two Dogmas of Empiricism", general criticisms of the concept of analyticity which may tell against the present notion. For the moment, though, the shored-up account faces more pressing problems.

Carnap's position is that with this one concession to Quine's anti-conventionalist arguments, he can proceed to define "analytic" in such a way that the truths of mathematics and of logical systems other than quantified first order logic do come out as analytic in the sense of being derived from conventions. Now, one problem with the account of "non-conventionalist" analyticity derived from the *Tractatus* is that the logic whose epistemic status is thereby supposedly assured is too weak to allow for the formalisation of anything more than primitive recursive arithmetic.⁴⁴ If the concept of analyticity was exhausted by the notion of analyticity in play in the *Tractatus*, then the moderate empiricists' commitment to (1) would require them to prescind from claiming a priori status for anything more than primitive recursive arithmetic. Carnap recognises this and proposes to use quantified first order logic as a metalanguage in which more powerful formal systems can be set up. The intention is to show that formulae provable by the linguistic rules of these more powerful systems will be counted as analytic in the sense of being derived from conventions. The linguistic framework equivalent to quantified first order logic Carnap calls Language I; the epistemic status of this language is supposed secured by the Tractarian considerations discussed above.⁴⁵ Language I forms the meta-language with which the linguistic rules of other linguistic frameworks can be set up. In line with the Principle of Tolerance⁴⁶, the linguistic frameworks set up using Language I as a metalanguage can be such as to

⁴⁴ *Ibid.*: 92. See also Quine 1966: 110.

⁴⁵ Carnap breaks with Wittgenstein in allowing that the syntactic rules of Language I are expressible within Language I. Wittgenstein, in the *Tractatus*, believed that this was impossible, holding that the rules could be "shown" but not "said", and that expressions of the syntactic rules of logic amounted to meaningless tautologies. Carnap, on the other hand, takes the work of Hilbert and Gödel to have refuted this view. See Carnap 1937: 281-84; Friedman in Aspray and Kitcher 1988: 85.

⁴⁶ See page 20.

Carnap's account of a priori knowledge

form sentences about any kind of phenomenon, physical, mathematical or otherwise; and we have already had occasion to note that this leaves Carnap with the difficulty of distinguishing, within the set of propositions expressed by sentences provable by the linguistic rules of a given language, the intuitively a priori propositions. We now turn to examine Carnap's attempted solution to this problem.

The problem of logical and physical rules

Following Carnap, we shall call a syntactic rule of a linguistic framework an *L-rule* just in case it is a rule for the formation and transformation of sentences of logic and mathematics; a *P-rule* will be a syntactic rule of a linguistic framework that is not an *L-rule*.⁴⁷ *P-rules* govern the formation and transformation of sentences referring to physical, that is, non-logico-mathematical, phenomena. It is intuitively plausible that there is such a distinction, but as it stands it has been drawn in terms of what terms refer to, that is, on semantic grounds. For Carnap, the problem of providing a moderate empiricist epistemology of a priori propositions is that of finding a syntactic criterion which will divide up the set of rules for a given linguistic framework in a way which matches the intuitive, semantic, distinction. Carnap believes that this criterion can be developed on the basis of a prior distinction between logical and descriptive vocabulary. This latter distinction also has intuitive plausibility, but while it remains at the level of intuition it is too vague and informal to do epistemological work.⁴⁸ It can, however, be clarified.

But if we reflect that all the connections between logico-mathematical terms are independent of extra-linguistic factors, such as, for instance, empirical observations, and that they must be solely and completely determined by the transformation rules of the language, we find the formally expressible distinguishing peculiarity of logical symbols and expressions to consist in the fact that each sentence constructed out of them is determinate.⁴⁹

⁴⁷ Carnap 1937: 180.

⁴⁸ *Ibid.*: 177.

⁴⁹ *Ibid.*

Carnap's account of a priori knowledge

A sentence of a linguistic framework is *determinate* just in case it is either a logical consequence of the empty set or a contradiction.⁵⁰ This appeal by Carnap to the concept of logical consequence, as opposed to the concept of syntactic consequence, is significant and will be returned to shortly. Given this definition of "determinacy", Carnap defines the set of logical expressions as the largest set *R* of atomic expressions such that every sentence which can be sub-divided solely into members of *R* is determinate.⁵¹

The distinction between L- and P-rules is intended to be defined on the basis of this definition. Note that it will not suffice to define a P-rule simply as any rule which contains descriptive vocabulary, since there can be rules with instances containing descriptive vocabulary which must be counted as logical. For example, where *D* is a descriptive predicate, the following expression is an instance of an L-rule⁵²:

$$(11) \quad Dx \rightarrow (\sim Dx \rightarrow Dx) .$$

To set up a definition of "L-rule" we need some new terminology.⁵³ A substitution for vocabulary within a sentence is *grammatically admissible* just in case vocabulary is substituted for by expressions of the same grammatical category; a substitution is *uniform* just in case the same value is assigned to equal variables. (4) is an L-rule because it is true, and remains true under any uniform grammatically acceptable substitution for the descriptive predicate *D*. In essence, then, an L-rule is a true sentence of a linguistic framework such that either (i) it contains only logical vocabulary or (ii) it contains descriptive vocabulary and remains true for any uniform grammatically admissible substitution for that descriptive vocabulary.⁵⁴

⁵⁰ *Ibid.*: 173.

⁵¹ *Ibid.*: 177-8. This definition is not satisfactory. $\exists x \exists y x \neq y$ meets Carnap's definition of a determinate sentence and yet it does not appear to be a matter of logic alone that there is more than one thing in the universe. Carnap does not appear to address this problem in *The Logical Syntax of Language*. (For discussion see Etchemendy 1990.)

⁵² Carnap 1937: 180-1.

⁵³ The terminology is not Carnap's own: it is drawn from Etchemendy 1990.

⁵⁴ *Ibid.*, and Carnap 1937: 180-1.

Carnap's account of a priori knowledge

By classing a sentence as analytic just in case it is an instance of an L-rule of a language, Carnap hopes to substantiate his intuition that all and only a priori propositions are analytic. The problem for this account lies, however, with his appeal to the concept of logical consequence in his account of the distinction between logical and descriptive vocabulary. As Carnap is well aware, Gödel's incompleteness theorem states that for a suitably strong formal system, the set of logical consequences of that system will be larger than the set of syntactic consequences - that is, there will be truths of that system which cannot be proved. It is only for first order predicate logic and systems weaker than this that the sets of syntactic and semantic consequences coincide.⁵⁵ The segregation of logical and descriptive vocabulary for a linguistic framework L (where L instantiates a formal system to which Gödel's incompleteness theorem applies; that is, it is one whose logical consequences outstrip its syntactic consequences) cannot therefore be established purely on the basis of the syntactic rules of that language. There is therefore no purely syntactic guarantee that the set of logico-mathematical vocabulary so defined will correspond to the intuitive set of logico-mathematical vocabulary. But Carnap cannot define a sentence as analytic just in case it is a logical consequence of L, since this fails to address the problem of distinguishing the L- and the P-consequences of the framework. Instead, in giving his general definition of analyticity⁵⁶, Carnap gives a recursive definition analogous to Tarski's recursive definition of truth.⁵⁷ Such a recursive definition of a property ϕ is intended to be able to establish, for any given expression p of a language L, whether ϕp . Now, for a language L, Tarski defines truth-in-L within a metalanguage L' composed of L augmented by a list of all the true atomic expressions of L.⁵⁸ Analogously, Carnap has to define analytic-in-L within a metalanguage L' which contains all the atomic expressions of L that are analytic. But we have access to such a metalanguage just in case *we have already determined* which atomic expressions of L are analytic.

⁵⁵ Hunter 1996.

⁵⁶ Carnap 1937: 110-12.

⁵⁷ Cf. "The Concept of Truth in Formalised Languages", reprinted in Tarski 1956.

⁵⁸ For simplicity, I assume that L is a propositional, not a quantified, language.

Carnap's account of a priori knowledge

In giving a recursive metalinguistic definition of "analytic-in-L" which will have the consequence that logic and mathematics are analytic, Carnap has to assume that logic and mathematics are analytic. Quine maintains that this decision is made arbitrarily, and that any other set of propositions could have been chosen for inclusion in the definition of "analytic". Given this decision, he writes,

...the thesis that logico-mathematical truth is syntactically specifiable becomes uninteresting. For, what it says is that logico-mathematical truth is specifiable in a notation consisting solely of [names of signs, an operator expressing concatenation of expressions], and the whole logico-mathematical vocabulary itself. But *this* thesis would hold equally if "logico-mathematical" were broadened[...] to include physics, economics, and anything else under the sun; Tarski's routine of truth definition would still carry through just as well. No special feature of logic and mathematics has been singled out after all.⁵⁹

This last sentence, though, is tendentious. Carnap, we might think, includes logic and mathematics in the metalinguistic definition precisely because he thinks they have a special status; namely, that they are a priori. To be sure, this assumption vitiates his argument that logic and mathematics are a priori because analytic, but it is important to be clear why this is so. In broad, schematic outline, we can see Carnap as wanting to prove

(12) The propositions of logic and mathematics are a priori.

He does this by assuming (1) and then giving an account of analyticity which has the consequence that

(13) The propositions of logic and mathematics are analytic.⁶⁰

But to prove (13) he has, in effect, to assume (12)! Someone who was well disposed to the existence of a priori knowledge might find it plausible to say that (12) is simply a datum and can well be assumed: the mistake, it might be thought, is to

⁵⁹ Quine 1966: 118.

⁶⁰ That is: they are expressed by sentences which are analytic.

Carnap's account of a priori knowledge

think that it needs justifying by appeal to (1) and (13). Carnap is in effect trying to give a reductive definition of the concept of apriority; his failure to provide one could be seen as evidence that the concept is simply independent of the concept of analyticity. Quine sees things differently, however. As we will see, Quine does not distinguish between apriority and analyticity; from his perspective, Carnap has simply assumed that logic and mathematics are analytic in order to prove that they are analytic.

Whatever interpretation we adopt, Carnap is unable to give any formal substance to his prior intuition that all and only logico-mathematical sentences are analytic; so his system of linguistic frameworks cannot help to explain analyticity in such a way as to demonstrate the truth of (1). In effect, in trying to formalise the distinction between logical and descriptive language, Carnap depends on the very intuition he is trying to formalise. Even granted a special epistemic status for language I, Carnap's attempt to provide an analytic theory of a priori knowledge fails.

CHAPTER THREE

Quine's critique of analyticity

Quine writes that the analytic theory goes "sturdily on" in the face of the doubts aired in the previous chapter.¹ This is unsurprising: since the moderate empiricists are committed to giving an analytic theory of a priori knowledge, they cannot take failure to give an account of analyticity that vindicates that theory as evidence that the theory itself is false. Quine's disparaging of the analytic theory might make it appear that he actually does reject it; this chapter casts doubt on this appearance.

A principled rejection of the analytic theory would involve, at the very least, the citing of counterexamples to (1)²; a more satisfying rejection would add to this an account of why an appeal to analyticity is inappropriate in giving an explanation of the nature of a priori knowledge. I argue in this chapter that Quine attempts only to do the latter. The claim he tries to establish is in effect the following: the severe problems with the concept of analyticity render it unsuitable for playing any explanatory role at all, and *a fortiori* unsuitable for explaining a priori knowledge. But there is no evidence that Quine rejects (1). Indeed, it seems he must accept it, for "Two Dogmas of Empiricism", the *locus classicus* of doubts about analyticity, begins with a critique of analyticity, and ends with an endorsement of what has come to be called "naturalised epistemology". Naturalised epistemology is often glossed as epistemology which rejects a priori knowledge; so, we might think that "Two Dogmas of Empiricism" contains an inference from the noninstantiation of analyticity to the noninstantiation of apriority; and the most natural way to do this is via the assumption of the left-to-right component of (1):

$$\forall p (p \text{ is a priori} \rightarrow p \text{ is analytic}) .$$

Since the right-to-left component, that is,

¹ Quine 1966: 119.

² The claimed equivalence between apriority and analyticity cited in chapter two.

Quine's critique of analyticity

$\forall p (p \text{ is analytic} \rightarrow p \text{ is a priori})$

is independently plausible, it would seem that Quine does not cite counterexamples to (1); and since it is plausible that the citing of such counterexamples is a necessary condition for a principled rejection of the analytic theory, he does not really reject the analytic theory. In fact acceptance of the theory is central to his epistemological project. If this is right, we might expect to see traces of the analytic theory in the theories of more recent philosophers who have reflected on a priori knowledge from a Quinean perspective. The following chapter will vindicate this expectation. The present chapter, however, is concerned with the role of the analytic theory in the development of Quine's own epistemology, and with showing that Quine's doubts about analyticity should not be construed as doubts about the analytic theory itself. Having done this, I go on to suggest that the widespread tendency to treat a priori knowledge as unrevisable knowledge³ arose because of Quine's own acceptance of the analytic theory; and I raise problems for the analytic theory which are independent of Quine's considerations. For now, however, these suggestions need bearing out with a close discussion of Quine's "Two Dogmas of Empiricism".

Quine's "Two Dogmas of Empiricism"

We will discuss "Two Dogmas of Empiricism" with two aims.

The first aim is to assess his critique of analyticity. This will involve three steps:

(i) *Identification of the notion of analyticity which Quine attacks.* We have already seen that the conventionalist account is not the only possible way of clarifying the concept of analyticity. We have seen two further options already: the attempt could be made to explain analyticity by appeal to the concept of necessity; or, following the ideas about the status of logic identified by Friedman in the *Tractatus* and the *Begriffsschrift*, an attempt

³ A belief is unrevisable, roughly, if there is no situation in which it is rational to give it up. A detailed definition is proposed in chapter five.

Quine's critique of analyticity

could be made to explain analyticity directly through appeal to the concept of meaning, without cashing out the latter concept in terms of conventions. Thus, there are a number of non-equivalent accounts of analyticity available; which of these, then, is Quine attacking? Or does he rather intend his considerations to apply to any account of analyticity whatsoever?

(ii) *Identification of the anti-analyticity thesis that he endorses.* As Grice and Strawson stress⁴, there are a number of ways of criticising a distinction:

It can be criticised for not being a sharp distinction (for admitting cases which do not fall clearly on either side of it; or on the ground that the terms in which it is customarily drawn are ambiguous (have more than one meaning); or on the ground that it is confused (the different meanings are habitually conflated).⁵

As they point out, Quine's aims in criticising the concept of analyticity are not of this sort. Adopting recent terminology, we can draw a distinction between a *nonfactualist* thesis and an *error thesis* about a putative class of entities, say, the class of α s.⁶ To adopt a nonfactualist thesis about α s is to maintain that the concept of an α is incoherent, and hence that expressions containing apparent references to α s are without a truth value. On an error thesis about α s, it is allowed that the concept of an α is coherent: a strong error thesis would then claim that, despite this, the property of being an α is necessarily uninstantiated; a weak error thesis claims that there is serious and widespread error in our beliefs about α s, such that the vast majority of our beliefs about α s are false. Given that Quine does not advance a modest critique of analyticity of the sort described by Grice and Strawson, which of these two more radical theses does he propound?

(iii) *Determination of whether Quine's attacks, suitably clarified, are cogent.*

There is particular concern that the arguments in sections 1 to 4, to the effect

⁴ Grice and Strawson 1956.

⁵ *Ibid.*: 141.

⁶ The terminology to be introduced drawn from Boghossian 1996/1997.

Quine's critique of analyticity

that the concept of analyticity admits of no non-circular definition, fail to establish any strong anti-analyticity thesis⁷; and there is also concern that a striking picture of language and meaning is advanced in the last two sections without significant argument or detail.

The second aim is to assess Quine's position on the analytic theory. Only once the above issues have been addressed will we be in a position to judge whether Quine's discussion in "Two Dogmas of Empiricism" bears directly on the analytic theory. We will see that Quine doesn't question the analytic theory itself.

The structure of "Two Dogmas of Empiricism" is apparently quite simple; this simplicity, as we have already suggested, masks underlying difficulties of interpretation. The two theses of empiricism Quine attacks are (i) that there is a privileged class of analytic truths and (ii) that verification conditions can be given for individual expressions of a language (this is the thesis of "atomic verificationism"). The arguments against (i) take up sections 1 to 4 of the paper. In these sections Quine examines and rejects a series of attempts to clarify the concept of analyticity. The apparent conclusion is that there is no way of giving a definition of "analytic" that does not appeal to concepts which are as much in question as the concept of analyticity itself. Section 5 considers the verification theory of meaning and argues briefly against the view that verification conditions can be given for expressions individually. Section 6 sketches a type of verificationist semantics, the adoption of which is apparently incompatible with maintaining the existence of a class of analytic truths; he then discusses the implications of adopting such a semantics for general philosophical and scientific issues.

However, once we try to give an interpretation of Quine's paper which goes beyond the superficial level things become more contentious. We therefore need to examine Quine's paper in more detail.

⁷ See for example, Grice and Strawson 1956, Dummett 1978, Priest 1979, Wright in Butterfield 1986.

Quine's critique of analyticity

Quine begins in section 1 by considering and rejecting two suggestions made by Kant. The first is that we can define "analytic" by appeal to the concept of self-contradiction. The idea would be that a sentence is analytic just in case it is self-contradictory to deny it. This, he concludes, simply shifts the burden of explanation onto the notion of self-contradictoriness, a notion in just as much need of clarification as the concept of analyticity.⁸ He finds equally unhelpful Kant's definition of an analytic judgement as one in which the predicate is contained in the subject. This is rejected as being applicable only to judgements and appealing to a notion of containment which is merely metaphorical. Quine draws from Kant, however, an expression of the intuitive notion of analyticity, namely that "a statement is analytic when it is true by virtue of meanings and independently of facts."⁹ The problem with this intuitive definition is that the concept of meaning it appeals to itself stands in need of clarification. Two distinct notions are typically conflated under the heading of "meaning". The first is reference (or "extension"), the reference of an expression being, in the standard formulation, that entity for which it stands. The second is sense (or "intension"), the sense being that feature of the meaning of the expression which determines its reference. Expressions with the same reference can have different sense: for example, the expressions "creature with a heart" and "creature with kidneys" are co-referring but differ in sense.

For reasons made clearer at a later stage in the paper¹⁰, and to which we will shortly return, the concept of analyticity cannot be clarified through appeal to a concept of meaning cashed out in terms of extension or reference. Analyticity, therefore, is a property at the level of sense. "Meaning", as used in the rest of Quine's paper, refers to sense; I will follow Quine's usage. Quine then expresses doubts that the theory of meaning can be understood as a theory of the properties of certain entities called "meanings": he asserts, without argument, that the theory of meaning is exhausted by the study of synonymous and analytic linguistic forms.

Once the theory of meaning is sharply distinguished from the theory of reference, it is a short step to recognizing as the primary business of the theory of

⁸ Quine 1953: 20.

⁹ *Ibid.*: 21.

¹⁰ *Ibid.*: 30-1.

Quine's critique of analyticity

meaning simply the synonymy of linguistic forms and the analyticity of statements; meanings themselves, as obscure intermediary entities, may well be abandoned.¹¹

Granting Quine this point, the lack of explanatory value of the intuitive definition of "analyticity" becomes clear: the appeal to "meaning" masks an appeal to the very concept which is to be explained.

Dropping discussion of the intuitive definition of analyticity, Quine distinguishes two classes of analytic sentences: there are the logical truths, which are those true sentences which preserve their truth through all uniform grammatically admissible substitutions for their nonlogical particles, and which Quine, in "Two Dogmas of Empiricism", treats as unproblematic¹²; and there are those sentences which are reducible to logical truths through intersubstitution of synonyms.¹³ For convenience, and not following Quine's own usage, I shall call sentences of the latter type "Frege-analytic".¹⁴

Frege-analytic expressions are synonymous with logical truths. Since Quine takes the analyticity of logical truths for granted, it seems we can give an account of Frege-analyticity just in case we can give an adequate account of synonymy. Quine accordingly turns, in sections 2 and 3 of the paper, to a discussion of attempts to define "synonymy".

Section 2 is given over to accounts of synonymy in terms of definition. Definitions, he points out, express prior relations of synonymy and hence cannot be appealed to in explaining synonymy. Much the same is true for Carnap's notion of explication. Explication does not restrict itself to reporting existing relations of synonymy, but attempts to improve the concept to be defined by refining its meaning; however, this process also appeals to, and therefore cannot explain, the concept of synonymy.

¹¹ *Ibid.*: 22.

¹² We pass over, as does Quine, the problem of defining what it is for an expression to be a logical particle. An early discussion of this problem can be found in Tarski's essay "On The Concept of Logical Consequence" (in Tarski 1956).

¹³ Quine 1953: 22.

¹⁴ After Frege's definition of "analytic" in *Foundations of Arithmetic* (Frege 1950). The terminology is used by Boghossian 1997.

Quine's critique of analyticity

Any word worth explicating has some contexts which, as wholes, are clear and precise enough to be useful; and the purpose of explication is to preserve the usage of these favoured contexts while sharpening the usage of other contexts. In order that a given definition be suitable for purposes of explication, therefore, what is required is not that the definiendum in its antecedent usage be synonymous with the definiens, but just that each of these favoured contexts of the definiendum, taken as a whole in its antecedent usage, be synonymous with the corresponding context of the definiens.¹⁵

We must note, however, that Quine admits of certain cases of definition in which a relation of synonymy is indeed set up: this is where a definiendum is introduced specifically for the purpose of being synonymous with the definiens. Quine calls this "the explicitly conventional introduction of novel notation for purposes of sheer abbreviation."¹⁶ He claims, however, that this is a special and rare case of definition and cannot help to explain synonymy generally.

In section 3 Quine considers whether synonymy can be defined by appeal to the concept of interchangeability. He begins by considering whether two expressions will be synonymous just in case they are interchangeable in all linguistic contexts without change of truth value (in Leibniz' phrase: "intersubstitutable *salva veritate*"). This idea requires refinement, however, since even for intuitive cases of synonym pairs such as "bachelor" and "unmarried man" there are contexts where intersubstitutability *salva veritate* fails. For example, the expression "she is a bachelor of arts" can be turned false (or ungrammatical) by substitution of "unmarried man" for "bachelor". Quine proposes to set aside such anomalous cases, provisionally by treating "bachelor of arts" and analogous expressions as a single word. We then modify our proposed definition of "synonymous" as follows:

- (i) Expressions α and β are synonymous just in case α and β are interchangeable in all linguistic contexts without change in truth value, except inside words.

¹⁵ Quine 1953: 25.

¹⁶ *Ibid.*: 26.

Quine's critique of analyticity

As Quine points out, the suggestion assumes that we have an account of what a word is, but by reducing the problem of synonymy to the problem of wordhood we would have made progress.

Quine now considers how interchangeability might be used to explain synonymy. We assume for the sake of example that "bachelor" and "unmarried man" are interchangeable within the constraints given in (i). Then, taking the obvious truth

(ii) Necessarily all and only bachelors are bachelors.

we can substitute for one occurrence of "bachelor" to generate:

(iii) Necessarily all and only bachelors are unmarried men.

This is, Quine asserts, equivalent to saying that the following sentence is analytic:

(iv) All and only bachelors are unmarried men

and if (iv) is analytic then 'bachelor' and 'unmarried man' are synonymous. Hence, it seems, interchangeability, suitably constrained, can be used to define "synonymy". However, the difficulty, Quine says, is that this argument makes free use of the concept of necessity. Quine's position in the 1951 paper is that the concept of necessity can itself only be explained by appeal to the concept of analyticity. A language containing the adverb "necessarily" is an intensional language, and Quine holds that intensional languages make sense just in case the concept of analyticity makes sense. To appeal to necessity in defining "synonymy" would presuppose that we have already made sense of analyticity, but the attempt

Quine's critique of analyticity

to define synonymy was embarked upon in order to make sense of analyticity.¹⁷ Interchangeability *salva veritate* cannot be used to define a notion of synonymy appropriate to giving an account of analyticity.

A purely extensional language is characterised by co-referring expressions of the language being interchangeable in all linguistic contexts without change of truth value. While a purely extensional language avoids the obscurities Quine associates with necessity, in such a language intersubstitutability except within words *salva veritate* does not provide a criterion for synonymy, as there is no way to distinguish statements that are true and well established, from those which follow from prior relationships of synonymy. For example, assuming that for ingrained social and biological reasons, marriage, and marriage alone, was the source of happiness: then "no bachelors are happy" and "no bachelors are married" are both true, but no concept of necessity or analyticity is available with which to draw a distinction between the types of truth which they possess.

These considerations exhaust Quine's discussion of the possibility of explaining analyticity via an account of synonymy. In section 4 he addresses the possibility that the concept of analyticity might be explained through appeal to the semantical rules of formalised languages. We have already discussed an early attempt on these lines, namely Carnap's attempt to define "analytic" using a syntactic criterion, and Quine's comments in the present section of the 1951 paper parallel the objections we have already seen raised against Carnap's project.

Quine considers two ways in which an appeal to the rules of artificial languages might help explain the concept of analyticity. First, he considers a language L whose rules just are the set of all analytic sentences of L; it seems, then, that for L we know precisely which sentences are analytic. But in the absence of an account of analyticity, Quine says, this procedure is empty: it attributes a property to the rules of L which has not yet been explained. It might be thought, however, that the set of rules of L at least can be used to frame a definition of "analytic-for-L":

¹⁷ *Ibid.*: 29-30. Quine's identification of the concepts of necessity and analyticity comes out clearly in his paper "Reference and Modality", reprinted in Quine 1953. See also his "Three Grades of Modal Involvement" in Quine 1966.

Quine's critique of analyticity

that is, a sentence *S* will be analytic-for-*L* just in case it is one of the rules of *L*. This procedure segregates a set of sentences of *L* and assigns a property to them, but does not help us pick out the analytic sentences of *other* languages, does not give us a general criterion for analyticity, and so, Quine thinks, does not help us understand the concept of analyticity at all.¹⁸

The second way in which artificial languages might be thought helpful is as follows. A semantical rule of a language can be thought of as implicitly defining a term within it by stipulating that some class of sentences containing that term are to be true (we discussed Carnap's early formulation of this idea above). We could, then, try classifying a sentence as analytic just in case it is true, and true in virtue of being assigned the value *true* by a semantical rule. But this procedure depends for its informativeness on our having some antecedent way of determining which are the semantical rules of a language. Not all sentences of a language can be semantical rules, on pain of making all true sentences of the language analytic. But the selection of a subset of true sentences of a language to be the semantical rules is done on the basis of convention and does not reflect any interesting property which can be used to define "analytic".

Semantical rules are distinguishable, apparently, only by the fact of appearing on a page under the heading "Semantical Rules"; and this heading is itself then meaningless.¹⁹

The present strategy, Quine suggests, is driven by a confusion of the role of postulates in setting up axiomatised formal systems. A formal system can be axiomatised by selecting as axioms certain theorems of that system which are then useful in deriving other results. The selected theorems are postulates, chosen on conventional and pragmatic grounds relative to some aim in setting up the system - perhaps, economy of formalisation. Those who try to define "analytic" by reference to semantical rules are treating semantical rules as postulates to the extent that postulates are chosen by convention, but fail to see that the postulates are not

¹⁸ Quine 1953: 33.

¹⁹ *Ibid.*: 34.

Quine's critique of analyticity

thereby made true by convention. Yet it is the latter property which would be required in order to arrive at a definition of "analytic" which corresponded to the intuitive notion.²⁰

These considerations exhaust Quine's attempts to find a definition of "analytic". He concludes by remarking on the naturalness of inferring from the double dependence of truth on language and fact, to the existence of a class of analytic truths which are true just in virtue of language. But this inference is mistaken:

For all its a priori reasonableness, a boundary between analytic and synthetic simply has not been drawn. That there is such a distinction to be drawn at all is an unempirical dogma of empiricists, a metaphysical article of faith.²¹

The direct discussion of analyticity over, Quine turns in section 5 to the verification theory of meaning, with which the second dogma is associated. The verification theory of meaning we have already mentioned in chapter one: roughly speaking, it is the view that the meaning of a sentence is closely associated with the set of observations which would confirm or infirm it, though a thoroughgoing definition is hard to come by. Quine discusses several modes in which the verification theory can operate. The earliest verificationist semantics, which Quine calls "radical reductionist", is that associated with Locke and Hume. On this view, the meaning of every term of a language is in some sense identified with a set of experiences, and the meaning of a compound expression is a function of these experiences. Such an account requires, in effect, every term to be counted as a *name* of certain experiences, and this view, though perhaps plausible in the case of proper names, is much less plausible when applied to such categories of terms as predicates, logical operators etc. A more plausible verification theory, developed by Frege and Bentham, attributes verification conditions only to entire sentences: that is, the sentence becomes the smallest unit of meaning. This we can call an

²⁰ *Ibid.*: 35. A more extended discussion of this confusion occurs in "Carnap and Logical Truth" in Quine 1966.

²¹ Quine 1953: 37.

Quine's critique of analyticity

"attenuated reductionist" thesis, since the strict term-by-term reduction of Locke and Hume has been abandoned. This view, Quine says, is *identical* with the view that there is a special class of analytic truths: the analytic truths will be those expressed by sentences which are confirmed by every experience.

The verification theory of meaning, which has been conspicuous in the literature from Peirce onward, is that the meaning of a statement is the method of empirically confirming or infirming it. An analytic statement is that limiting case which is confirmed no matter what.²²

The dogma of reductionism, even in its attenuated form, is intimately connected with the other dogma - that there is a cleavage between the analytic and the synthetic. We have found ourselves led, indeed, from the latter problem to the former through the verification theory of meaning. More directly, the one dogma clearly supports the other in this way: as long as it is taken to be significant in general to speak of the confirmation and infirmation of a statement, it seems significant to speak also of a limiting kind of statement which is vacuously confirmed, *ipso facto*, come what may; and such a statement is analytic.

The two dogmas are, indeed, at root identical...²³

Quine proposes that the move from radical to attenuated reductionism does not go far enough: the attenuated account still takes too narrow a view of the units of meaning. Sentences, Quine suggests, only have meaning as part of a body of sentences²⁴; and on the next page he goes further, stating that the body of sentences within which individual sentences have meaning is identical with the totality of scientific theory.²⁵ Quine's position, that only in the context of the entirety of science does a sentence have meaning, can be termed "holistic verificationism", and is returned to in section 6. The grade of reductionism intermediate between "radical reductionist" and "holistic" can usefully be termed "atomic verificationism".²⁶

In the sixth and final section of the paper, Quine fills out some details of the holistic semantic theory and draws some consequences from it. Quine's theory of

²² *Ibid.*: 37.

²³ *Ibid.*: 41.

²⁴ *Ibid.*: 41.

²⁵ *Ibid.*: 42.

²⁶ See Føllesdal in Guttenplan (ed.) 1975.

Quine's critique of analyticity

language is commonly explained through the introduction of the metaphor of a "web of belief", though in fact this metaphor is not used in the 1951 paper. Quine does appeal to two different metaphors - that of the totality of knowledge as being a "man-made fabric which impinges on experience only along its edges", and of total science as being a "field of force whose boundary conditions are experience" - but this is in the context of giving an account of the revision of beliefs which is for the most part uncontroversial. Quine describes the properties of the system of beliefs as follows.

A conflict with experience at the periphery occasions readjustments in the interior of the field. Truth values have to be redistributed over some of our statements. Reevaluation of some statements entails reevaluation of others, because of their logical interconnections...²⁷

This is simply a picture of belief-revision induced by experience, and constrained by requirements of consistency within the belief-set. The picture is augmented, however, in three significant ways.

First, and without argument, Quine adds that the logical connections between sentences can also be revised on the basis of experience. The quotation begun above continues:

Reevaluation of some statements entails reevaluation of others, because of their logical interconnections - the logical laws being in turn simply certain further statements of the system, certain further elements of the field. Having reevaluated one statement we must reevaluate some others, which may be statements logically connected with the first or may be the statements of the logical connections themselves.²⁸

Second: the core and the periphery of the system of beliefs²⁹ are to be distinguished by reference to the degree to which specific experiences bear on the revision of a given belief. The more peripheral a belief, the fewer the experiences

²⁷ Quine 1953: 42

²⁸ *Ibid.*

²⁹ Quine talks of the revision of *statements*, rather than beliefs. However, I think the idea of the revision of a belief is more intuitive than the idea of the revision of a statement, and I have recast my commentary on Quine accordingly.



Quine's critique of analyticity

which could lead us to revise it; more central beliefs can be revised in response to indefinitely many different experiences.³⁰ However, this distinction is limited in the following ways. Quine asserts (a) that any belief can be held true come what may: that is, even beliefs close to the periphery can be retained by corresponding adjustments deeper in the system; (b) that no belief is immune to revision: that is, even beliefs in logical laws can be revised under certain circumstances.

Third, Quine suggests that these observations are meant to apply, not just to the system of knowledge and belief, but to language and meaning.

If this view is right, it is misleading to speak of the empirical content of an individual statement - especially if it is a statement at all remote from the experiential boundary of the field.³¹

This is a significant addition. For while the remarks about the holistic nature of the relation between theory and evidence are uncontentious, the adoption of a holistic verificationism has profound and counterintuitive consequences.

Quine draws two consequences from this account of knowledge and meaning. First, from points (a) and (b) it follows that that there is no sense in distinguishing a special class of analytic sentences.

...it becomes folly to seek a boundary between synthetic statements, which hold contingently on experience, and analytic statements, which hold come what may.³²

Second, his account of knowledge and meaning entails that the distinction between metaphysical frameworks and scientific theory, enshrined by Carnap's distinction between internal and external linguistic frameworks, must be abandoned.³³ Postulation of metaphysical entities, such as physical objects, forces or abstract entities, is as much a part of the general project of fitting experience into a conceptual framework as the creation of scientific theories. What are normally

³⁰ *Ibid.*: 43.

³¹ *Ibid.*

³² *Ibid.*

³³ *Ibid.*: 45.

Quine's critique of analyticity

thought of as metaphysical postulates differ from scientific theories only in point of how ingrained they are in our ways of thought. It follows that metaphysical claims are open to empirical disconfirmation in the same way as scientific theories; again, the difference lies only in the relative immunity to counterevidence which we afford metaphysical claims, due to their centrality within the system of language and knowledge.³⁴

Interpretation and discussion

With a detailed account of "Two Dogmas of Empiricism" now before us, it is time to attempt the task of evaluation. The first major task we set ourselves was to assess Quine's critique of analyticity; the first step in doing this was to identify the concept of analyticity which Quine attacks. In section 1, Quine states explicitly that he is interested in sentences which are reducible to logical truths through intersubstitution of synonyms - that is, the Frege-analytic sentences. Since the concept of synonymy is central to the notion of Frege-analyticity, the fact that Quine takes up the bulk of sections 2 and 3 with a critique of synonymy bears out his claim to be centrally interested in this notion of analyticity. However, Quine also makes comments which bear on three other notions of analyticity.

First, in reducing the study of meaning to the study of synonymous and analytic linguistic forms, Quine may be blocking an appeal to the sort of Tractarian notion of analyticity appealed to by Carnap to secure the special epistemic status of logic in the face of Quine's "Truth by Convention". This suggestion has to be hedged since, as we remarked, it is not altogether clear what this notion actually is. But if Quine is right about the study of meaning being reducible in this way, he is indirectly blocking accounts of analyticity which seek to deploy the notion of meaning "unreduced".³⁵

³⁴ *Ibid.*: 44-5.

³⁵ Contemporary theories of intentional content could perhaps be thought of as giving an account of meaning "unreduced", in the present sense. If this is right, then Quine's equation of the study of meaning with the study of analyticity and synonymy may pass over possibilities for accounting for analyticity. Jerry Katz could perhaps be seen as offering an account on these lines; certainly Boghossian 1997 seeks to give such an account. See chapter nine.

Quine's critique of analyticity

Second, in section 4 Quine puts the notion of synonymy aside, and discusses attempts to define "analytic" by appeal to semantical rules. Quine argues that the explication of "analyticity" in terms of semantical rules would be informative only if the concept of analyticity were already understood. Here, then, Quine also discusses a non-Fregean notion of analyticity; essentially, he is discussing accounts related to Carnap's conventionalist account, discussed in the previous chapter.

Third, section 3, though primarily concerned with the possibility of giving an account of synonymy in terms of intersubstitutibility, does contain a lemma to the effect that we can understand what it is for a language to contain the adverb "necessarily" just in case we can understand the concept of analyticity. This follows from Quine's assumption, expressed in "Two Dogmas of Empiricism" and elsewhere, that a sentence of the form "necessarily, p " is true if and only if p is analytic.³⁶ Quine, therefore, concerns himself tangentially with an account of analyticity in terms of necessity. From a contemporary perspective, his claim that the concept of necessity and the concept of analyticity are equivalent has very little plausibility. If Quine were right, there would indeed be little point in appealing to necessity in defining "analytic"; however, work in modal metaphysics done since 1951 has shown that the concept of necessity does afford of a characterisation independent of the concept of analyticity.³⁷ As already mentioned, for space reasons I intend where possible to skirt metaphysical questions in this thesis. However, it is worth mentioning that although Quine is wrong to think that the concept of necessity can only be defined through appeal to analyticity, it turns out that his claim that intersubstitutibility *salva veritate* cannot be appealed to in defining synonymy retains plausibility. For it seems unlikely that, for all p and q , if p is necessarily equivalent³⁸ to q then p and q are synonymous. For example, Goldbach's conjecture is presumably either necessarily true or necessarily false, but we do not know which. It follows that one of these two propositions is true:

$(1+1=2 \leftrightarrow \text{Goldbach's conjecture is true})$

³⁶ For a general account of Quine's scepticism about necessity, see his paper "Necessary Truth", broadcast by the Voice of America in 1963, and reprinted in Quine 1966.

³⁷ The classic work being Kripke 1980.

³⁸ That is: if $(p \leftrightarrow q)$.

Quine's critique of analyticity

$(1+1=2 \leftrightarrow \text{Goldbach's conjecture is false})$.

If necessary equivalence was indeed sufficient for synonymy, we must say that one of the pairs $(1+1=2, \text{Goldbach's conjecture is true})$, $(1+1=2, \text{Goldbach's conjecture is false})$ is a synonym pair: but we do not know which. And it stretches credibility to claim that there can be synonym pairs which we are unable to recognise as such. True, some logicians have been willing to claim this, and say that mathematical discovery consists in the "unfolding" of hidden synonymy relations³⁹, but this idea has never seemed a plausible account of mathematical advance. Commitment to this view seems sustained mainly by antecedent commitment to the analytic theory of a priori knowledge. Suffice to say that it is indeed difficult to give an account of synonymy in terms of necessity, though for reasons not considered by Quine.

Hence, although Quine's chief interest lies with Frege-analyticity, his arguments also bear on other interpretations of the concept. However, his concern with these remains tangential to the main theme of sections 1 to 4 which is that the notion of synonymy is no better understood than the notion of analyticity, and that the problems with the latter transfer over to the former. We might feel that very compelling grounds would have to be offered before we would be willing to do something as counterintuitive as adopt a non-factualist or error thesis about synonymy. This reaction, I think, is appropriate, and we will discuss it shortly.

The second step towards assessing Quine's critique of analyticity requires that we determine the nature of Quine's anti-analyticity thesis. Two of Quine's comments seem to indicate that the thesis he advances is not a non-factualist thesis; that is, he is not trying to show that the concept of analyticity is incoherent.

First, in section 1 he seems to allow that logical truths are analytic, and then sets them on one side for the purposes of the discussion in the rest of the paper.⁴⁰

³⁹ For example, Ayer 1936.

⁴⁰ Quine 1953: 22.

Quine's critique of analyticity

Second, in section 2 he allows that there is one case of definition which really does set up a relation of synonymy between definiens and definiendum, namely the "explicitly conventional introduction of novel notation for purposes of sheer abbreviation".⁴¹

From these two claims it would appear that Quine does allow for the existence of isolated cases of analyticity and synonymy. If this appearance is upheld, and if it is also the case that Quine endorses a nonfactualist thesis about the concept of analyticity and synonymy, then it might seem that Quine's position is unintelligible.⁴² One way to avoid attributing unintelligibility to Quine would be to construe him as advancing, not a nonfactualist thesis about analyticity, but a weak error thesis. To hold a weak error thesis about analyticity is to hold that, in general, our beliefs about analyticity are misguided and full of error, and that the vast majority of propositions naïvely thought to be analytic truths turn out to be synthetic truths. It is consistent with maintaining a weak error thesis that we allow there to be isolated cases of analyticity and synonymy. One way of motivating the attribution of an error thesis to Quine, then, might be to reflect on the two comments just cited.

However, we might well doubt whether Quine's admissions in those two comments are indeed inconsistent with his maintaining a nonfactualist thesis. With regards the first comment, on a closer inspection of the text it appears that Quine does *not* allow that logical truths are analytic.⁴³ What he actually says is that the substitutional definition of logical truth renders the notion of logical truth unproblematic, and also that logical truths are analytic "by popular acclaim". This does not amount to an admission that logical truths are analytic truths.⁴⁴

Likewise, with regards the second comment, we might doubt whether allowing for cases of abbreviative definition really amounts to acceptance of a distinction between synonymous and nonsynonymous linguistic forms. What Quine actually says is that the idea of abbreviative definition is unproblematic, but that reflection on such cases does not provide a way of characterising "synonymy"

⁴¹ *Ibid.*: 25-6.

⁴² Cf. Grice and Strawson 1956: 152. See also Hunter 1995.

⁴³ Quine 1953: 22-3.

⁴⁴ *Contra*, in particular, Hunter 1995.

Quine's critique of analyticity

in general.⁴⁵ To have a clear and coherent distinction we require more than just a single class of cases which are alleged to fall on one or the other side of it. We also need a procedure for deciding in general where relevant cases fall with respect to it. The claim that cases of abbreviative definition are cases of synonymy does not help us in developing such a procedure. Hence, we should not take Quine's comments as being in tension with a nonfactualist thesis about synonymy.

Although the preceding considerations do not establish that Quine endorses an error thesis about analyticity, it is arguable on different grounds that it is appropriate to attribute such a position to him. For we must recognise that a nonfactualist thesis about analyticity also induces a nonfactualist thesis about syntheticity. That is, since "analytic" and "synthetic" are correlative terms, there cannot be no facts about where one term applies without there being no facts about where the other term applies. But a synthetic claim is, roughly, one that is open to experiential confirmation and disconfirmation; and we have seen that Quine holds that all claims are so open, even mathematical, metaphysical or logical claims. Thus Quine's conclusion in "Two Dogmas of Empiricism" appears to be precisely the claim that all truths are synthetic truths. This position *is* inconsistent with maintaining a nonfactualist thesis about analyticity. Quine's anti-analyticity thesis must therefore be taken to be an error thesis. Further, we should take him to be maintaining a *strong* error thesis, to the effect that the properties of analyticity and synonymy are necessarily uninstantiated: for Quine's holistic semantics rules out the possibility of any sentence which is confirmed come-what-may.

Having determined what notion of analyticity Quine is criticising and what the substance of his criticism is, we turn to evaluating his arguments. Many authors have found Quine's arguments in the 1951 paper unconvincing.⁴⁶ Dummett in particular finds it strange that Quine should spend four sections arguing that there is no way of giving an informative definition of "analytic", and then turn to a discussion of a semantic theory, namely atomic verificationism, on which the

⁴⁵ Quine 1953: 25-6.

⁴⁶ See for example Dummett 1978, Priest 1979, Wright in Butterfield 1986.

Quine's critique of analyticity

concept of analyticity can be given a clear definition, *viz.*: as a statement which is verified by any and every experience ("vacuously confirmed").⁴⁷ Accordingly, Dummett suggests that "Two Dogmas of Empiricism" divides into two parts. The first part, consisting in sections 1 through 4, propounds the thesis that "the concepts of analyticity and syntheticity are spurious, on the ground that it is impossible to give non-circular definitions of the related terms."⁴⁸ If this were all there was to the argument it would be weak, for there is no obvious reason why we should object to the adoption of terms which are defined in such a way.⁴⁹ Not all concepts must be reducible to be legitimate, on pain of infinite regress, and Quine gives us no reason to think that the concept of analyticity is not such an atomic concept. Dummett suggests that Quine's real doubts about analyticity arise in the second part of the paper, composed of sections 5 and 6, where analyticity is linked with atomic verificationism, and atomic verificationism is itself challenged. The principal contribution to the philosophy of language made by "Two Dogmas of Empiricism", Dummett thinks, is the presentation of holistic verificationism.⁵⁰

Dummett's interpretation makes "Two Dogmas of Empiricism" out to be unconvincing. It makes the arguments in the first part seem not only weak, but irrelevant to the work done in the second part. Moreover, the argument in the second part also appears to be weak, even nonexistent, for as we have seen, holistic verificationism is put forward with very little argument. The immediate motivation for this holism seems to be nothing more than the adduction of an empiricist tradition of enlarging the units of meaning: from words, to entire sentences, and hence, Quine suggests, to the whole of knowledge. And in section 6 the more detailed account of holism in language turns out to be an uncontentious view of holism in *epistemology*, that is, holism in the way theory and evidence interact, which is then augmented without argument with three nontrivial theses (that the logical laws are just more statements in the system of knowledge; that any statement can be held true come what may and no statement is immune to revision; and that the epistemological picture applies to language and meaning). Dummett

⁴⁷ Dummett 1978: 375.

⁴⁸ *Ibid.*

⁴⁹ See also Grice and Strawson 1956.

⁵⁰ Dummett 1978: 376.

Quine's critique of analyticity

concludes that Quine's 1951 account of holistic verificationism is too sketchy to be evaluated and turns to Quine's other works, primarily *Word and Object*, for a more informative account.⁵¹

But Dummett's interpretation may be uncharitable to Quine. If we recall Quine's insistence that the two dogmas are at root identical, we may be able to reconstrue the structure of "Two Dogmas of Empiricism" in such a way that the argument therein gains in plausibility.

If the two dogmas are identical, then an argument against the one is also an argument against the other; hence Dummett's claim that holistic verificationism is presented without an argument to support it may be unfair. Quine takes it that the adoption of holistic verificationism is forced by the rejection of atomic verificationism; so if Quine is right that acceptance of atomic verificationism is identical with acceptance of the existence of analytic truths, arguments against the analytic/synthetic distinction will tell against atomic verificationism and so in favour of Quine's holism. Problems raised for the analytic/synthetic distinction transfer over to atomic verificationism via the thought that it would be odd, given that atomic verificationism makes it so natural to talk of sentences that are vacuously confirmed, that we should have no coherent intuitive account of what these sentences are. But sections 1 to 4 indicate that no account of the intuitive notion is forthcoming: hence, we might conclude, perhaps the notion of vacuous confirmation is chimerical, and there is something wrong with the semantics - atomic verificationism - which allows for such things.

The flaw, of course, is that 1 to 4 do not establish that no account of the intuitive notion of analyticity is forthcoming: as noted, Quine's argument that the concepts of synonymy, necessity and analyticity can only be defined in terms of each other fails to establish negative conclusions about them: in particular, it leaves open the possibility that analyticity is an atomic concept. However, the proposed identity between the two dogmas may be appealed to once more, to lend plausibility to Quine's arguments in sections 1 to 4.

⁵¹ *Ibid.*: 377ff.

Quine's critique of analyticity

In sections 5 and 6 we see Quine claiming that no feature internal to a sentence determines its confirmational status: depending on the context, any belief can be treated as vacuously confirmed, and with a shift in context it can lose that status. It is clear, though, that the classical notion of analyticity requires a truth to be confirmed in *all* contexts if it is to be analytic. The claim that any belief can be taken as not-vacuously-confirmed trumps the claim that any belief can be taken to be vacuously-confirmed; the element of decision Quine perceives in the determination of the confirmational status of beliefs should be taken as showing that properly speaking, no belief is vacuously confirmed. If Quine is right, and the property of being vacuously confirmed is indeed uninstantiated, then it is at least odd to claim that the concept of analyticity is atomic. Atomic concepts, we might think, have a particularly fundamental role in our conceptual scheme, and refer to properties of particular significance - goodness, truth, causation, etc. It is hard to see how an uninstantiated property could come to have such significance for us; hence if analyticity is uninstantiated it is hard to see how it could be an atomic concept.

This suggestion is, of course, inconclusive. No strong reason has been given to think that our histories could not have furnished us with a spurious atomic concept. The point in raising it, though, is to suggest that Quine's arguments against the two dogmas might gain in plausibility through being construed as standing in a relation of mutual support. The argument that there are problems with the definition of "analytic" supports Quine's claim that atomic verificationism is false; atomic verificationism is taken to be identical with the doctrine that some sentences are vacuously confirmed; and the suggestion that there are no sentences that are vacuously confirmed (that is, that all beliefs are revisable) supports the claim that there are problems with the definition of "analytic".

But, while the relation of mutual support envisaged may allow two individually inconclusive arguments to lend plausibility to each other, it will not help against the stronger claim that the arguments are not merely inconclusive but fundamentally flawed. And there are serious problems with Quine's arguments.

Quine's critique of analyticity

Our reading Quine as maintaining that the concept of analyticity cannot be defined, and is not atomic, is in tension with our decision to take him as defending an error thesis about analyticity. Arguments that there are serious problems with the very definition of the notion of a thing, α , tend to establish that expressions containing apparent reference to α are actually meaningless: but this is an extreme non-factualist thesis.

The natural way of responding to this worry is to read Quine as maintaining a strong error thesis about synonymy based on his defence of holism. If individual expressions do not have a unique meaning then there is no sense in comparing individual expressions with respect to their meaning. However, the very idea of synonymy is not thereby taken to be incoherent, since the notion of meaning itself is not taken to be incoherent (just misapplied). Indeed, we might introduce instances of synonymy by stipulating that two expressions are synonymous just in case they make the same contribution to the meaning of the total system.⁵² This explains, I think, Quine's proposal to allow rare cases of synonymy through explicitly conventional introduction of new terminology.

However, our intuitions about synonymy are deeply ingrained; very compelling reasons must be offered if we are to give them up; and there is reason to think that Quine's holism does not provide reason enough.

First, the project of giving a worked out holistic semantics faces serious technical problems. Holism entails that no sentence of a subject's idiolect has a meaning of its own: its meaning depends on all the other beliefs the subject has, on all the other inferences that she is making. Hence, no (verbal, somatic, etc.) expression by the subject has a content independent of the rest of the speaker's beliefs. This generates severe problems for understanding how communication is possible: communication presumably depends on recognition of shared meanings, and it is hard to see how any meanings can be shared if holism is true. Rather, the contents of expressions will be radically idiosyncratic, and will be interpreted

⁵² This suggestion is close to Grice and Strawson's proposal that we can take two expressions to be synonymous just in case any experiences which, on certain assumptions about the truth values of the rest of the system of expressions, confirm or infirm the one, also, on the same assumptions, confirm or infirm the other. Grice and Strawson 1956: 156.

Quine's critique of analyticity

idiosyncratically by the hearer.⁵³ There are of course responses to this problem (foremost the work of Davidson), though there is no consensus on whether these responses are adequate. Suffice to say here, however, that the filling out of the holistic framework is a nontrivial task.

Second, and more generally, Dummett has claimed that the adoption of a holistic semantics amounts to an admission that the project of giving a theory of meaning cannot be successfully completed. A theory of meaning aims to present an algorithm which would, for any expression, generate another expression giving the meaning of the first.⁵⁴ If no expression has a meaning of its own, this project is a non-starter. In many of his works, including "Two Dogmas of Empiricism", Quine argues for a general scepticism about meaning which would, if upheld, count against the possibility of giving a theory of meaning; but we will not here discuss Quine's arguments, or the controversy they have provoked.⁵⁵ The consensus is that Quine's attempt to establish scepticism about meaning is unconvincing: the very fact that the arguments Quine offers would, if sound, establish such scepticism, has tended to seem reason enough to reject them.⁵⁶

Third, and analogously to the previous point, Boghossian has suggested that the clearest reason to reject holism is precisely the scepticism about synonymy it engenders. This suggestion has great force: our intuitions about synonymy run deep, and as we have just seen, holistic verificationism cannot claim an equivalent status. I conclude that Quine's attempt to establish an error thesis about Frege-analyticity via considerations of semantic holism, fails. Quine gives us no overwhelming reason to reject synonymy; hence, it seems that if logic is analytic, there is indeed a class of further sentences which are also analytic.

But our rejection of Quine's anti-analyticity arguments leaves two main issues untouched. The first of these concerns Quine's legacy: his attack on

⁵³ Cf.: Fodor and LePore 1992.

⁵⁴ Cf.: Miller 1998.

⁵⁵ There is an enormous body of literature on this point: see in particular Wright's recent survey paper, Wright 1997.

⁵⁶ This point is made by Boghossian 1997.

Quine's critique of analyticity

analyticity has been extremely influential, and features of the way in which he argues recur in even the most recent discussions of a priori knowledge. The second issue concerns the status of the analytic theory of a priori knowledge. While this theory stands, the vicissitudes of the concept of analyticity infect the concept of apriority. If we wish to discuss these separately, the analytic theory must be rejected.

Quine's legacy

Quine's influence on the development of theories of a priori knowledge can be discussed under two headings: first, the unrevisability of a priori knowledge; second, the promotion of naturalism in philosophy and epistemology.

Regarding the first point, Quine's account of analyticity in terms of vacuous confirmation nicely encapsulates a crucial epistemic feature of analytic truths: being confirmed by every experience, no experience can count against their truth. Hence, where *S* believes in a proposition *p* which is expressed by an analytic sentence, there is no evidential situation in which it is rational for *S* to give up her belief in *p*. That is to say, roughly, that beliefs in analytic propositions are unrevisable.⁵⁷

Given the analytic theory, the unrevisability of beliefs in analytic truths will entail the unrevisability of beliefs in a priori truths. From here it is a short step to the claim that a priori knowledge is unrevisable knowledge. This must be considered Quine's major legacy to theories of the a priori.

The second aspect of Quine's legacy follows closely on the first. The view that a priori knowledge is unrevisable, coupled with the arguments in 5 and 6 of "Two Dogmas of Empiricism", to the effect that all knowledge is revisable, engenders scepticism about a priori knowledge. Such scepticism has been taken to be central to - and even as the defining feature of - epistemological naturalism. If Quine is right, then the possibility of giving a nonsceptical naturalistic account of a priori knowledge is ruled out. We will discuss this scepticism in detail in the next chapter.

⁵⁷ The notion of unrevisability is discussed in depth in chapter five.

Quine's critique of analyticity

Evaluating the analytic theory

We suggested earlier that there are two ways of offering a principled rejection of the analytic theory of a priori knowledge; the minimum requirement would involve the citing of counterexamples to the proposed biconditional relationship between propositional analyticity and apriority (proposition (1) of chapter two). A more satisfying rejection would add to this an account of why the appeal to analyticity is incapable of explaining a priori knowledge. Quine, we should note, does only the second of these: on his view, analyticity is incapable of explaining *anything*. He does not, though, offer counterexamples to (1). That is, Quine's argument against analyticity does not meet the minimum necessary condition for it to be an argument that the analytic theory of a priori knowledge is false. So he does not really reject the analytic theory at all.

If we wish to challenge the analyticity theory, our first task is to see if counterexamples are forthcoming. I suggest that they are. In the first place, as many anti-empiricist authors have pointed out, there are a series of propositions such as

Red is a colour

$\forall x \forall y (\text{If } x \text{ is later than } y \text{ then } y \text{ is earlier than } x)$

No surface can be entirely red and blue simultaneously,

which, though intuitively known a priori, do not seem to be analytic.

Examples such as these, however, are controversial; some have claimed these propositions to be analytic, and the unclarity about what the concept of analyticity actually involves tends to sustain this controversy.⁵⁸ Boghossian, for

⁵⁸ See for example discussion between Boghossian 1996 and Harman 1996.

Quine's critique of analyticity

example, suggests that such propositions can be considered implicit definitions of certain component terms.⁵⁹

However, I do not think that we have to rest much on these disputed examples. Our recent considerations regarding unrevisability offer us a more promising route. Consider an extremely complicated logical truth p , which S proves only after great mental effort. For simplicity, let us say that p is a theorem of a propositional logic system with set of axioms PL . Having proved p , S knows p a priori. According to the analytic theory p is an analytic truth, and hence S 's belief in p cannot be given up without change of language. Some time later, let us suppose, S reads in an authoritative journal of cognitive psychology a paper which claims that the majority of people who try to construct proofs of the sort that S used in proving p succumb to a subtle logical trick and end up proving the opposite of the truth. Reading this, and being unable at the time to rule out the possibility that she has succumbed to that trick, but realising that it is likely that she has in fact succumbed to it, S reasonably revises her belief in p : she now believes $\neg p$. Let us call the evidence that engenders this belief revision, e . p , since it was revised, is not analytic; nevertheless, before the reception of e , it was known a priori.

The moderate empiricist might respond to this that p , since revisable, was not known a priori to start with. This response is not obviously wrong: a priori knowledge is experience-independent knowledge, and there is clearly a sense in which revisable knowledge is not knowledge which is independent of experience. Chapter five will be centrally concerned with whether a priori knowledge is unrevisable knowledge. We will eventually conclude that there is a crucial sense in which a priori knowledge is not unrevisable: if this conclusion is right, the analytic theory is blocked.

If the moderate empiricists want to deal with this case without assuming that a priori knowledge is unrevisable knowledge, they will have to allow that p was known a priori, but that the revision of S 's belief in p engenders a change in S 's language. The reasoning behind this is that, since beliefs in analytic truths are unrevisable without change of language, and S 's a priori knowledge of p was

⁵⁹ Boghossian 1997; see also chapter nine.

Quine's critique of analyticity

revised, S's language must have changed along with change in her beliefs with regard to p . Let's call the language she speaks before receipt of e , L_1 , and the language she is supposed to speak after receiving e , L_2 . Clearly this shift from L_1 to L_2 must be imperceptible to S, since she would not agree that her language had changed, though the idea of an imperceptible change in language does not seem to be too problematic. There are more serious problems, however. We can stipulate that, before and after receiving e , S is fully aware of, and has entirely true beliefs about, the basic truth-tabular definitions of the logical connectives. Since the truth tables give the meaning of the connectives of propositional logic, it seems plausible that if S's language changes in point of logic, this change must show up in her beliefs about the meanings of the logical connectives. But since her true beliefs about the truth-tabular definitions of the connectives remain unchanged, it seems extremely implausible to hold that S speaks a different language after revising her belief in p . A more plausible interpretation of S's predicament is that in revising her belief in p upon receipt of e , and so coming to believe $\neg p$, S's belief set gains an inconsistency. S believes in all the axioms of the system of propositional logic she is using, but believes $\neg p$, even though $\{PL\} \vdash p$. But it is hard to see how the moderate empiricist can accommodate this view. First: the inconsistency induced in S's belief set by her revision of her belief in p is an inconsistency couched in language L_1 . It is simply not clear whether this inconsistency holds in language L_2 , since we do not know enough about it. But the very fact that it is not clear whether the inconsistency holds is a strike against the moderate empiricist's treatment of the case of revision of a priori knowledge. Second, S does not notice the change in language the empiricist must say is forced by the reception of e , and so does not notice that she no longer shares a language with her former self. But if communication depends on the possession of a shared language by those communicating, it is hard to see how S can understand her past utterances, or those of others who speak L_1 . An imperceptible proliferation of languages is the consequence of adopting the empiricist account of apriority, and this just seems to be a very implausible view of language. We should conclude that S's knowledge of

Quine's critique of analyticity

p is a priori, but that p is not an analytic truth. If this is right, the analytic theory of a priori knowledge is false.

CHAPTER FOUR

Naturalism and a priori knowledge

In concluding the previous chapter we argued that the analytic theory of a priori knowledge fails for reasons independent of Quine's doubts about analyticity. Quine, indeed, has nothing to say about the analytic theory itself, and, it seems, simply assumes it to be true. Further, we argued that Quine fails to establish a strong negative thesis about analyticity, with the consequence that even if the analytic theory were true, he would fail to establish any strong conclusions about a priori knowledge itself. However, we also noted two ways in which Quine's work has influenced subsequent thinking about a priori knowledge. First, he promotes the view that a priori knowledge is unrevisable knowledge; second, he promotes a naturalistic approach to philosophy in general, and epistemology in particular, which many have taken to rule out a priori knowledge.

If Quine is right that naturalised epistemology rules out a priori knowledge then the project of naturalising a priori knowledge is a non-starter. So, if we are to develop a plausible naturalistic theory of a priori knowledge we need to find an appropriate account of naturalism within which to frame it. The primary aim of this chapter is to sketch such a framework. Having done this, we will look at an influential theory of a priori knowledge which is naturalistic in the sense that we endorse: this is the theory offered by Philip Kitcher in *The Nature of Mathematical Knowledge*.

Overview of theories of the a priori

It is standard in the literature to assume that any possible theory of a priori knowledge will fall into one of precisely three categories.¹ These are:

- (i) analytic theories
- (ii) rationalist theories
- (iii) scepticism.

¹ See for example Bonjour's (1993, 1998) survey of accounts of a priori knowledge.

Naturalism and a priori knowledge

The first of these is associated with the moderate empiricists, and has already been rejected. In the previous chapter it was argued that although all analytic truths may well be knowable a priori, not all a priori knowledge is of analytic truths.

Rationalists adopt the view that a priori knowledge cannot be explained using the same conceptual apparatus as employed in explaining knowledge in general. Accordingly, they postulate a new epistemic capacity, through the exercise of which we have a priori knowledge.² Rationalists are often charged with appeal to obscurities; it is claimed that the new epistemic capacity, often explained through analogy with perception, cannot be taken beyond the level of analogy.³ These charges may be unfair, since there are some well worked-out rationalist positions available⁴; however, I do not discuss the rationalist response in this thesis. My concern is with whether there is an interesting notion of a priori knowledge which is explicable within the framework of scientific and epistemological concepts we currently have available: accordingly, accounts which require significant augmentation of that framework will be ignored.⁵

Having rejected both the analytic theory and rationalism, the standard view as set out above would require that we adopt scepticism: that is, that we deny that there is any a priori knowledge. A cluster of theories which are thought to give rise to scepticism standardly go under the term "naturalism". Scepticism about a priori knowledge is associated with naturalism because of the influence of Quine: as suggested in the previous chapter, and as we will see in more detail below, Quine's own naturalism is inseparable from his doubts about analyticity and apriority. However, Quine's version of naturalism is not uniformly shared: one does not have to be a Quinean to be a naturalist.

² See for example BonJour 1993/1998; Russell 1912: 105 (discussed in Creath 1993: 143).

³ For criticisms see for example Harman 1996; Boghossian 1997; Creath 1993.

⁴ See Brown 1993, and BonJour 1993, 1998, for example.

⁵ This terminology is not uniformly adopted: Peacocke's "moderate rationalist" account of a priori knowledge explicitly eschews appeal to unexplained properties of intuition and the like, and tries to explain a priori knowledge through reference to the conditions which a subject must meet if they are to count as possessing a concept. On my classification, Peacocke's theory would count as naturalist. It is structurally similar to Boghossian's account of a priori knowledge in terms of implicit definition, and I discuss Boghossian's account in chapter nine.

Naturalism and a priori knowledge

A recent paper by Devitt provides an exceptionally blunt illustration of the perceived connection between naturalism and scepticism about a priori knowledge.⁶ Devitt's paper is specifically concerned to defend the view that a priori knowledge is incompatible with naturalism. Outlining what he means by "naturalism", Devitt writes:

The naturalism in question is an epistemological doctrine that I take from Quine: there is only one way of knowing, the empirical way that is the basis of science (whatever that way may be). So I reject "a priori knowledge". I do not give a detailed account of my rejection but I do give two reasons[...]: Briefly, first, [...] we lack a strong motivation for thinking that mathematics and logic are immune from empirical revision; and, second, the idea of a priori knowledge is deeply obscure, as the history of failed attempts to explain it show.⁷

The first sentence of this quotation, taken in isolation, suggests that Devitt simply defines naturalism in such a way that it will exclude a priori knowledge. But clearly, such a strategy will not provide a strong motivation for scepticism about the a priori: for, faced with it, the apriorist will just seek a redefinition of "naturalism" which removes this scepticism, and the debate threatens to become purely verbal. If the question of whether naturalism is inconsistent with a priori knowledge is to become substantive, we need two things: (i) an account of naturalism which has *prima facie* plausibility; (ii) demonstration that that account of naturalism poses problems for a priori knowledge. Devitt realises this, for he goes on to give two reasons to be sceptical about a priori knowledge. The first, essentially, is that we have reason to think that all knowledge is revisable; the second is that we have reason to think that the concept of a priori knowledge is obscure. Neither of these reasons is convincing as it stands. Clearly, if the first reason is to establish scepticism, it needs supplementing with an argument that a priori knowledge is unrevisable knowledge; while the support adduced for the second reason is an induction from past failures of attempts to define a priori knowledge, and does not establish an anti-a priori thesis that is specifically naturalistic. Further, if by

⁶ Devitt 1998. Kitcher (1992) also builds rejection of a priori knowledge into his definition of "naturalism".

⁷ Devitt: 45.

Naturalism and a priori knowledge

"obscure" Devitt means "incoherent", his anti-a priori thesis faces a problem analogous to that which we have argued afflicts the non-factualist version of the anti-analyticity thesis. Since "a priori" and "a posteriori" are correlative terms, an argument that the concept of a priori knowledge is incoherent would also be an argument that the concept of a posteriori knowledge is incoherent. But Devitt does not want to claim that the concept of a posteriori knowledge is incoherent: in the quotation above he states explicitly that he takes *all* knowledge to be a posteriori knowledge (assuming only that Devitt follows common practice in taking "empirical knowledge" to be a synonym for "a posteriori knowledge"). The challenge that I think Devitt poses is not to show that the concept of a priori knowledge is coherent, but rather, granting that the notion of experience-independent knowledge can be made acceptably clear, to explain how there ever could be such knowledge. This problem has been posed most forcefully by Benacerraf, whose arguments we will look at shortly. Kitcher's account of a priori knowledge, which will concern us closely, proceeds by giving a precise definition of a priori knowledge and then arguing that none of our knowledge in fact meets the standards embodied in that definition.

The task we have to undertake, then, is to find a formulation of a naturalist position which we seem to have good reason to accept; then we will see if either of Devitt's two reasons to reject a priori knowledge can, within the context of that account, motivate scepticism about a priori knowledge.

We proceed by looking more closely at naturalism. Devitt distinguishes two varieties of naturalism.⁸ The first, which he calls "epistemological naturalism", is the view that all knowledge has the same epistemic status as scientific knowledge, which is supposed to be a posteriori knowledge. The second, which he calls "metaphysical naturalism", he claims is equivalent to physicalism, which we can gloss as the view that the ontological claims made by physics constitute a complete and true account of the ontology of the world. We will discuss each of these in turn, to see if they meet both of the following requirements: (i) they are interesting and

⁸ *Ibid.*: 46.

Naturalism and a priori knowledge

plausible philosophical positions, and (ii) they pose problems for a priori knowledge.

Epistemological naturalism

Since Devitt explicitly links his scepticism about a priori knowledge with epistemological naturalism, it is natural to begin with the latter; and since Devitt calls this version of naturalism Quinean, it is natural to discuss Quine's own account of naturalism.

Quine develops his view of naturalised epistemology through contrast with epistemology as traditionally construed. According to the traditional account, the project of epistemology is to reconstruct knowledge, to distinguish between that which we know and that which we only think we know, to expose false beliefs and confer certainty on what remains.⁹ This project is, in Kitcher's phrase, *meliorative*.¹⁰ In distinguishing the certain from the doubtful, traditional epistemology was thought to enable us to improve our epistemic practices, revealing flaws and helping us to overcome them. It often seems that philosophy as a whole is identified with this project, which is inherited from Descartes. Transposed into the context of moderate empiricism, epistemology gains a distinctive methodology: Descartes' project is to be carried out by exhibiting the connection between knowledge and experience; a claim to know *p* is to be evaluated through determining whether the experiences that would verify *p* had actually occurred.

Quine sees traditional epistemology as a combination of two projects: the reduction of complex concepts to simpler components, and the proof of some knowledge claims by appeal to others.¹¹ In conjunction, these two projects yield a strategy of analysing complex sentences into simple components which have intuitively clear truth conditions; the clarity of the latter is then supposed to transfer up the analytical chains and confer equivalent epistemic status on the complex sentences. This process is supposed to "validate" scientific knowledge. In effect,

⁹ See also discussion of empiricism in chapter one.

¹⁰ Kitcher 1992: 64.

¹¹ Quine 1969: 69-70.

Naturalism and a priori knowledge

claims to scientific knowledge, which are considered contentious, are to be reduced to uncontentious claims to nonscientific knowledge, concerning sense experience and analytic truths.

The two projects, however, have met with at best limited success. Conceptual investigations have yielded insight into the foundations of mathematics and the relations between logic and mathematics, but have provided sparse insight into aspects of a posteriori knowledge. The conceptual reduction of physical objects to sets of experiences proposed by Hume, for example, does not greatly improve the epistemic status of knowledge claims concerning physical objects. As regards this project Quine writes:

What then of [...] the justification of our knowledge of truths about nature? Here, Hume despaired. By his identification of bodies with impressions he did succeed in construing some singular statements about bodies as indubitable truths, yes; as truths about impressions, directly known. But general statements, also singular statements about the future, gained no increment of certainty by being construed as about impressions.¹²

Quine takes his semantic holism to present a general problem for traditional epistemology. He sees traditional epistemology as proceeding by taking individual knowledge claims and seeing if they are based on the experiences which would verify them. But the thesis of semantic holism is, as we saw in the last chapter, just that sentences of a language do not have verification conditions individually. If this is true, and if traditional epistemology is inextricably wedded to atomic verificationism as Quine assumes, then traditional epistemology is an attempt to carry out the impossible.

Faced with this, epistemology must at least be extensively modified, perhaps even abandoned. Quine's proposal is that the Cartesian project to reconstruct knowledge should be abandoned, but that epistemology should be retained, construed now as the project of understanding the relationship between experience and knowledge. Epistemology, in fact, just is a branch of empirical psychology on Quine's view. As he writes:

¹² *Ibid.*: 71-2.

Naturalism and a priori knowledge

Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. This human subject is accorded a certain experimentally controlled input - certain patterns of irradiation in assorted frequencies, for instance - and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. The relation between the meagre input and the torrential output is a relation that we are prompted to study for somewhat the same reasons that always prompted epistemology; namely, in order to see how evidence is related to theory, and in what ways one's theory of nature transcends any available evidence.¹³

This reconstrual, Quine thinks, engenders a rethinking of the limits we impose on what we may appeal to in looking for this understanding: in particular, we should not disallow an appeal to psychology as a possible way of understanding how the construction of knowledge from experience really proceeds.

As Quine points out, from the point of view of traditional epistemology such an appeal to scientific knowledge in the process of carrying out the epistemological project was thought to induce circularity. The project of traditional epistemology was to vindicate claims to knowledge and until that project has been carried out, scientific knowledge should be considered *sub judice*. To appeal to such knowledge in the process of vindicating our claims to possess it seems to presuppose that which we are trying to establish: namely, that our claims to this knowledge are valid. We are supposed to be reducing scientific knowledge claims to non-scientific knowledge; this project is defeated if scientific knowledge is also allowed to be present in the reduction base. Thus traditional epistemology, as carried out by the moderate empiricist, seems to be a uniquely philosophical enterprise: it is meant as an assessment of knowledge claims that by its very nature rules out contributions from the sciences. Philosophy, in fact, is considered to be an autonomous discipline, exclusively concerned with analytic truths.¹⁴ The hazy chain of reasoning underpinning this view is something like the following. The project of philosophy is to vindicate knowledge claims; therefore, it can only avail itself of knowledge which does not need vindicating, that is, a priori knowledge. The analytic theory,

¹³ *Ibid.*: 82-3.

¹⁴ Cf. Kitcher 1992.

Naturalism and a priori knowledge

assumed by the traditional epistemologists, then engenders the view that philosophy is concerned only with the relations between concepts. The study of philosophy is, it is supposed, nothing more than the study of language. Following the title of Descartes' revolutionary prolegomena to the mechanical philosophy, his *Meditations on First Philosophy*, traditional epistemology is often called "first philosophy", and the slogan "no first philosophy" characterises much thinking post-Quine which takes the failure of the traditional project, and the circularity involved in Quine's naturalistic reconstrual, simply to rule out epistemology.¹⁵ Quine, however, holds that the bar against this form of circularity falls simultaneously with the project of traditional epistemology. Traditional epistemology has to disallow appeal to science precisely because it seeks a reconstruction of scientific knowledge in non-scientific (and so supposedly more perspicuous) terms; for Quine, to drop the goal of reconstruction is also to drop the view that epistemology requires reductive analysis of scientific knowledge.

If the epistemologist's goal is validation of the grounds of empirical science, he defeats his purpose by using psychology or other empirical science in the validation. However, such scruples against circularity have little point once we have stopped dreaming of deducing science from observations. If we are out simply to understand the link between observation and science, we are well advised to use any available information, including that provided by the very science whose link with observation we are seeking to understand.¹⁶

But the dropping of the bar against circularity is not the only consequence stemming from Quine's view of naturalised epistemology. In the previous quotations we see no hint that the meliorative aspect of traditional epistemology is to be retained. Quine marginalises this aspect when he says that the goal of seeing how evidence relates to theory, central to naturalised epistemology, was also the fundamental motivation of traditional epistemology. For Quine, naturalised epistemology drops the normative aspect of epistemology: it settles for scientific description of the way knowledge stems from sensory inputs; it does not seek to

¹⁵ Quine cites Polányi, Kuhn and Russell Hanson in this tradition (Quine 1969: 87). More recently, the position has been championed by Rorty (see especially Rorty 1980).

¹⁶ Quine 1969: 75-6.

Naturalism and a priori knowledge

determine whether the evidence really does justify our knowledge claims, and it does not seek ways to improve our epistemic practices.

Quine's naturalism, then, can be said to have two aspects: denial of the autonomy of philosophy; and abandonment of the normative project of traditional epistemology in favour of description.

The rejection of analyticity is central to Quine's argument that his version of naturalism should be adopted. As we saw in the last chapter, he believes that the rejection of analyticity forces a move to a holistic verificationist semantics, and it is this semantics which is meant to undermine traditional epistemology. But, by the analytic theory, the rejection of analyticity is also a rejection of a priori knowledge. Quine's naturalism and his scepticism about a priori knowledge are therefore connected through the common root of the rejection of analyticity.

However, it is precisely at this point that the conclusions from the last chapter are relevant: we have argued both that Quine fails to make a convincing case that analyticity should be rejected, and, moreover, that the analytic theory of a priori knowledge is false. If these arguments are sound, Quine's case for naturalising epistemology is undermined. But, though Quine's arguments for naturalism may fail, much of his insight into the nature of philosophy can still stand. Naturalism as a position can be extricated from the arguments Quine uses to motivate it.

For, in the first place, traditional epistemology as reconstructed by the moderate empiricists, from which the autonomy of philosophy is meant to flow, is hopeless flawed. It faces technical problems - no serious progress had been made in assigning verification conditions to individual knowledge claims¹⁷; conceptual problems - the central concept of a "sense datum", an entity whose essence is to be an appearance, is deeply suspect¹⁸; and motivational problems - it is simply not clear

¹⁷ See Quine's comments on Carnap's attempts in this direction, Quine 1966: 74-7.

¹⁸ See especially Sellars' influential essay "Empiricism and the Philosophy of Mind", reprinted in Sellars 1963.

Naturalism and a priori knowledge

that knowledge claims need vindicating in the way traditional epistemology assumes.¹⁹

More generally, though, the claim that philosophy is autonomous is undermined precisely by the failure of the analytic theory of a priori knowledge. The autonomy of philosophy is supposed to be secured by claiming philosophy to be concerned exclusively with analytic truths. But this stipulation about the remit of philosophy is not meant to be arbitrary: the proponents of this view, taking for granted the analytic theory, presumably assume that by making this stipulation they secure the traditional view that philosophy is concerned with a priori knowledge. But with the failure of the analytic theory, proponents of the autonomy of philosophy must face a dilemma. The set of analytic truths and the set of a priori truths are distinct, and a decision must be made about which of these sets philosophy is chiefly concerned with: either philosophy is fundamentally about analyticity, or philosophy is fundamentally about a priori truth. The first option preserves autonomy by trivialising philosophy: for philosophy remains the study of linguistic truths but there are no longer good grounds to think that the study of linguistic truths is the study of important truths, such as those known a priori. The second option preserves the importance of philosophy by undermining its autonomy: philosophy is concerned, as traditionally thought, with a priori knowledge, but it has nowhere been shown that science too is not also concerned with such knowledge. This leads naturally to the view that no firm distinction can be drawn between science and philosophy on epistemological grounds, that is, on the grounds of what sort of knowledge they are concerned with. Only a general distinction can be made: science is typically more concerned with a posteriori knowledge than is philosophy. A key aspect of Quine's naturalism therefore stands: philosophy and science are engaged in the same project, that of understanding the world, and philosophy should not isolate itself from the sciences in pursuing this project. Kitcher writes that this impulse to deny the autonomy of philosophy is what unifies the diverse naturalist positions.²⁰ I take this view of the

¹⁹ To drop the project of validating knowledge claims is not to drop the meliorative project entirely: we can seek to evaluate and improve our epistemic practices without seeking a total reduction of knowledge to sense data, logic and set theory.

²⁰ Kitcher 1992: 55.

Naturalism and a priori knowledge

relationship between philosophy and science to be central to the plausible naturalist position relative to which we will discuss a priori knowledge.

What of the other aspect of Quine's naturalism, that of rejecting the meliorative, normative aspect of traditional epistemology in favour of a project of pure description? This has not been widely thought to be compelling.²¹ A second unifying feature of much naturalism post-Quine has been to retain the meliorative project, while seeking reductions of normative and epistemic concepts to scientifically acceptable terms.²² What counts as a "scientifically acceptable term" will be addressed shortly. The second aspect of our plausible naturalist position is, then, that it retain the meliorative project of traditional empiricism.

We now turn to see whether the sketch of naturalism we have before us poses problems for a priori knowledge. The first of Devitt's reasons for rejecting a priori knowledge seems most relevant here. Devitt's charge, we recall, was that a priori knowledge is unrevisable knowledge and that there is no room in a naturalised epistemology for unrevisable knowledge. Now, the latter claim is plausible. The view that we should never treat any scientific belief as immune to possible refutation is associated most strongly with C.S. Peirce. Rescher describes this position as follows.

We would like to think of our science as "money in the bank" - as something safe, solid, and reliable. Unfortunately, however, history militates against this comfortable view of our scientific theorizing. In science, new knowledge does not just supplement but generally upsets our knowledge-in-hand. We must come to terms with the fact that - at any rate, at the scientific level of generality and precision - *each* of our accepted beliefs *may* turn out to be false, and *many* of our accepted beliefs *will* turn out to be false.²³

²¹ Cf. Foley 1994. As Foley points out, Quine himself retracts this aspect of naturalism in later writings.

²² Kim 1988 sees this as a project to show how the normative, and specifically the epistemic, supervenes on the non-normative.

²³ Rescher 1984: 83.

Naturalism and a priori knowledge

Rescher suggests that this *Peircean fallibilism*²⁴ applies to science but not to our ordinary common sense beliefs.²⁵ Intuitively, this is strange: we might naturally think that the procedure of scientific discovery produces results which are less rather than more vulnerable to error. In any case, it is plausible that fallibilism also applies in areas such as common sense beliefs, and even in mathematics and logic. Errors can be made in calculation, deceptions perpetrated, and the number of beliefs which seem immune to such difficulties seems extremely small.²⁶

But what of the assumption that unrevisability is the defining feature of a priori knowledge? As suggested in the previous chapter, this view is a hangover from the analytic theory of a priori knowledge, and we have not yet seen any independent support for it. We should note that the fact that epistemological naturalism renders it likely that all knowledge is revisable indicates a constraint on the theory of a priori knowledge which we will go on to develop: the theory should not have as a consequence that a priori knowledge is unrevisable knowledge. However, Devitt's argument ultimately fails because we will find no good reason to hold that unrevisability is the mark of a priori knowledge.²⁷ We will discuss it in detail with respect to Kitcher's definition of a priori knowledge.

Metaphysical naturalism

It appears that no immediate problems for a priori knowledge are posed by the epistemological version of naturalism; so we turn to the metaphysical version. Devitt, we recall, identifies metaphysical naturalism with physicalism, the view that physics provides a true and complete account of ontology. However, physicalism should really be considered a special case of metaphysical naturalism in general, which we can take as the view that philosophy should look to science to answer the question of what exists: that is, we should include in ontology only things which are

²⁴ To be distinguished from the sort of fallibilism of warrant to be discussed in chapters seven and eight.

²⁵ See Rescher's entry on "Fallibilism" in Honderich 1995.

²⁶ Chapter five contains an argument for a related form of fallibilism based on the details of the naturalist epistemological framework.

²⁷ This is also urged by Rey in his companion piece to Devitt's paper (Rey 1998); this view of a priori knowledge as revisable is increasingly popular, as the following chapters will demonstrate.

Naturalism and a priori knowledge

scientifically acceptable. For the physicalist, the scope of the scientifically acceptable is limited to the ontology of physics. Physicalism is controversial, however, and unduly restrictive for our purposes. We can perhaps view the choices available to us in characterising the "scientifically acceptable" ontology as governed by two parameters. The first parameter concerns the degree of intertheoretic reduction we require. A strict stance as regards the first parameter would restrict the scientifically acceptable ontology to the ontology of physics; to be acceptable, an entity must be shown to be metaphysically reducible to some combination of entities countenanced by physics. A more permissive stance would allow unreduced appeal to objects, properties or relations of higher level sciences, ascending through chemistry, biology, psychology, sociology, etc. The second parameter concerns whether we admit only currently accepted entities, or whether we allow entities that seem potentially scientifically acceptable. A strict stance here would require that only properties, objects and relations that appear in current scientific theories may be appealed to; a permissive stance would allow entities to count as scientifically acceptable if they seem to have potential to be brought into scientific theory. For example, the cranial pulse is recognised by osteopathy, but not by orthodox medicine; but there seems to be no fundamental barrier to prevent the phenomenon from becoming generally taken up in mainstream medical science.

Devitt's physicalism seems to take a strict stance on both parameters: he restricts the scope of the scientifically acceptable to the ontology of actual physics; Papineau's physicalism takes a strict stance on the first and a more permissive stance on the second, being couched in terms of a "future completed physics".²⁸ Fodor would take a permissive stance on the first parameter, allowing unreduced appeal to the entities of the special sciences.²⁹ It is not easy to generalise about what a permissive stance on the second parameter will allow, but one point seems clear: long standing metaphysical commitments debar appeal to normativity in science. Hence the naturalist programme is one of eliminating normative concepts, or showing that they are reducible to concepts referring to non-normative properties. In particular, epistemic concepts cannot be appealed to unreduced by naturalists.

²⁸ Papineau 1993.

²⁹ See also Botterill and Carruthers 1999.

Naturalism and a priori knowledge

Devitt's second reason for rejecting a priori knowledge was that the concept was obscure, and we interpreted him as claiming not that the concept of a priori knowledge is incoherent, but that it is difficult to give a plausible account of how a priori knowledge possible. The perspective of metaphysical naturalism has been thought to bolster this view. The concept of knowledge, being epistemic, seems banned from appearing unreduced in a naturalistic theory; it will therefore need analysing into naturalistically acceptable terms if it is to be legitimately employed. Now, many have thought that a plausible naturalistic account of knowledge will be a causal one; it will invoke an epistemically significant causal relationship between the knower and the thing known. As Benacerraf argued in his 1973 paper "Mathematical Truth", this view causes problems in accounting for a priori knowledge.

Benacerraf's intention in this paper is to highlight the tension between giving a perspicuous account of mathematical knowledge and adopting a semantics of mathematical knowledge claims that is continuous with the semantics of empirical knowledge claims. We have, he suggests, good reason to prefer an account of the semantics of mathematical knowledge claims that will ascribe to both

(1) There are at least three large cities older than New York

and

(2) There are at least three perfect numbers greater than 17

the logical form of:

(3) There are at least three FGs that bear R to a .³⁰

The truth value of (1), which uncontroversially has the form of (3), can be determined through standard application of the Tarskian recursive definition. The

³⁰ Benacerraf 1973: 663.

Naturalism and a priori knowledge

Tarskian definition explains truth in terms of concepts of reference and satisfaction. Application of the definition requires that we interpret the predicates, objects and relations of (3) realistically, that is, that we take there to be genuine objects, bearing the relevant predicates, which stand to each other in the specified relation. To assimilate (2) to (3), then, requires that we take a realistic view of mathematical propositions, interpreting these propositions as being about certain mathematical objects, predicates and relations. Benacerraf notes that such a realistic view of mathematics, which he calls the "standard view", is unacceptable to some. The alternative view, which he calls "combinatorial", is exemplified by Carnap's attempt to distinguish empirical from logic-mathematical propositions on syntactic grounds.

The leading idea of combinatorial views is that of assigning truth values to arithmetic sentences on the basis of certain (usually proof-theoretic) syntactic facts about them. Often, truth is defined as (formal) derivability from certain axioms [...] In any event, in such cases truth is conspicuously not explained in terms of reference, denotation or satisfaction. The "truth" predicate is syntactically defined.³¹

Combinatorial views do not treat (2) as having the logical form given in (3). The motivation behind combinatorial views, Benacerraf suggests, is epistemological: the standard view, though it has the advantage of treating mathematical discourse with the same semantic apparatus as ordinary empirical propositions, adopts a metaphysics of the objects of mathematical discourse which leaves it mysterious as to how we have mathematical knowledge. For we have no account of what epistemically significant relations can hold between cognitive subjects and the objects of mathematics as construed realistically.³² However, combinatorial views suffer from not giving an account of how the possession of a privileged syntactic property bears on the truth value of a proposition; the failure of Carnap's attempt to set up an analytic theory of a priori knowledge exemplifies the problems with this approach. Benacerraf takes it that the only viable account of truth is Tarski's; therefore, there is a presumption in favour of an account of mathematical truth that

³¹ *Ibid.*: 665.

³² *Ibid.*: 673.

Naturalism and a priori knowledge

treats it realistically.³³ But this leaves us with the problem of accounting for mathematical knowledge.

Benacerraf's observations constitute a significant problem for a naturalistic account of a priori knowledge. We will see that in the development of theories of knowledge, there was a move away from causal theories to reliabilist theories, and that these latter have better potential to solve, or at least side-step, the problem that Benacerraf raises.

The results of this chapter so far are as follows. The naturalist framework within which we will seek to accommodate a priori knowledge can be encapsulated in three claims: first, that philosophy is not autonomous from science; second, that philosophical theories must be developed within a scientifically acceptable ontological framework. (Though we have not committed ourselves as to what counts as scientifically acceptable, the crucial point is that epistemological concepts be analysed in non-normative terms.) Third, that the "meliorative" aspect of traditional epistemology can be retained: that is, that it will not be necessary to abandon any vestiges of the Cartesian project in favour of a radical Quinean project of pure description. Following Kitcher, we will call this position "traditional naturalism". Kitcher would augment our three claims with a fourth: that there is no a priori knowledge.³⁴ But, as we have argued, it is uninteresting simply to build this into our definition of naturalism. A more meaningful approach is to see whether the three claims themselves entail that there is no a priori knowledge; and we have seen two areas in which we could develop arguments that they do indeed entail this. First, it could be argued that a priori knowledge is unrevisable knowledge, and that there is no unrevisable knowledge; second, it could be argued that our general epistemology leaves no room for experience-independent knowledge. We will discuss these in the following chapters.

³³ *Ibid.*: 670.

³⁴ Cf. Kitcher 1992: 74-6. Kitcher gives four claims which he takes to characterise what he calls "traditional" naturalism. The framework I have suggested above is essentially the same as Kitcher's minus his requirement that there is no a priori knowledge. I suggest that this requirement needs to be demonstrated on the basis of the other claims: it is not axiomatic.

Naturalism and a priori knowledge

The remainder of this chapter discusses Kitcher's attempt to give a naturalistic theory of a priori knowledge. Before turning to this, however, we should note that some authors seem to have taken the loss of autonomy by philosophy as raising problems. Quine is often cited as arguing that philosophy, naturalistically construed, becomes in effect "high level", abstract, theoretical science, and this is thought to be problematic. But we should note that by the same token, science must be seen as "low level", specific, natural philosophy: this seems to be a view of the relationship between science and philosophy with a longer and more respectable history than the view which takes philosophy to be an autonomous discipline, concerned exclusively with conceptual analysis.

The blurring of the distinction between science and philosophy does not entail that the sorts of knowledge claims made by these two camps cannot in general be distinguished. It is right, I believe, to take philosophy, broadly construed, to be dealing with questions susceptible to answer a priori. Where the empiricists erred was in assuming that the only theory of the a priori could be an analytic one; this error having been accepted, the assumption became entrenched that the subject matters of philosophy and science are also distinct. The perceived difficulties with seeing philosophy as continuous with natural science stem from specifically empiricist considerations which we have no good reason to accept.

Naturalised epistemology and a priori knowledge

The contemporary project of providing a naturalised account of a priori knowledge was inaugurated by Philip Kitcher in *The Nature of Mathematical Knowledge*.³⁵ Kitcher's account is particularly interesting for our purposes, since he gives a clear definition of a priori knowledge within a naturalistic framework. Clearly, then, he does not find the concept of a priori knowledge obscure or incoherent: however, his account has the consequence that we in fact possess very little a priori knowledge.

Kitcher divides the problem into two sub-problems: the first involves giving a naturalistic theory of knowledge in general; the second involves giving an account

³⁵ Kitcher 1983.

Naturalism and a priori knowledge

of the notion of *experience-independence*. The solutions to these two sub-problems are then to be combined to give an account of experience-independent knowledge: that is, a priori knowledge. Further, Kitcher takes it that these two sub-problems can be answered independently of each other: in fact, as we will see, he proposes to leave the first to "general epistemology", taking only the outlines of a solution as a framework within which to develop a solution to the second.

The central problem faced by theories of knowledge in general is that of explaining the difference between knowledge and mere true belief. Following Kitcher, we can use the term *warrant* to refer to whatever property makes that difference. Warrant is the property *X* which solves the equation:

$$\text{Knowledge} = \text{True Belief} + X.$$

A warranted true belief is, by definition, an item of knowledge. Naturalistic (Kitcher also calls them "psychologistic") theories of knowledge assume that for a belief to possess warrant is a matter of it having a certain *aetiology*, that is, a certain causal history. As Kitcher puts it:

[On] a psychologistic account... the question of whether a person's true belief counts as knowledge depends on whether the presence of the state of true belief can be explained in an appropriate fashion. The difference between an item of knowledge and mere true belief turns on the factors which produced the belief - thus the issue revolves around the way in which a particular mental state was generated.³⁶

Certain *belief forming processes*, then, are such that, under certain circumstances, they warrant the beliefs they produce. This suggestion, that warrant is a function of *aetiology*, leads Kitcher to refine the general equation for knowledge given above, by building in the claim that warrant is conferred by appropriate belief forming processes. The result, which is as follows, he calls a "simple normal form for knowledge"

³⁶ *Ibid.*: 13.

Naturalism and a priori knowledge

- (1) S knows p iff p is true, S believes p , and S's belief in p was produced by a process which is a warrant for it.³⁷

Strictly speaking, this "normal form" will not do, since we must recognise the following points. First, that a belief which was appropriately produced, and so is warranted at the time of production, can lose that warrant at a later stage; second, that a belief that was inappropriately produced, and so is not warranted, can gain warrant at a later stage (perhaps as new evidence comes in, and new belief-forming processes become available). Kitcher's account, which locates warrant purely with context of production, faces counterexamples based around these two points. However, it is a simple matter to augment Kitcher's "normal form" so that these counterexamples are removed: let a belief be warranted at a time just in case it is appropriately *sustained* at that time, let an appropriate process which sustains a belief be a warrant for that belief, and let production be a special case of sustaining.³⁸ Since this problem with Kitcher's account is superficial, and since it is common for naturalistic theories of warrant to couch their accounts simply in terms of production, I will not refer to this point again.

In line with his strategy of dividing the problem of a priori knowledge into two sub-problems, Kitcher does not give any detailed suggestions as to what properties a process must have in order to be "appropriate"; this he seeks to leave to general epistemology.³⁹ Having defined knowledge as (roughly speaking) warranted true belief, Kitcher goes on to use the concept of warrant as an undefined primitive in giving an account of a priori warrant. Arguably, Kitcher's most significant methodological advance is that of reducing the problem of explaining a priori knowledge to the problem of explaining a priori warrant. Since a priori knowledge is a species of knowledge in general, intuitively described as

³⁷ *Ibid.*: 17.

³⁸ Cf.: Summerfield 1991: 94 n.31.

³⁹ Although subsequent authors have tended to assume that Kitcher meant his account to be reliabilist, it is worth noting that Kitcher does not explicitly say this in his 1983 presentation. There is, however, strong indirect evidence that he did have reliabilism in mind: see for example his endorsement of Goldman's epistemology at Kitcher 1983: 18 n.6. In later work Kitcher has made this assumption explicit (Kitcher 1992: 75-6).

Naturalism and a priori knowledge

"experience-independent" knowledge, it is plausible to think that the a priori warranting processes will be a species of warranting processes in general: they will be the experience-independent processes. Kitcher's hope - and this hope is shared by authors who have developed his ideas - is that an account of experience-independence can be given independently of an account of warrant in general.

The task of defining a priori knowledge therefore becomes the task of defining a priori warrant: that is, of setting the standards which warrants must meet if they are to be experience-independent. And so we come to the second of Kitcher's sub-problems, that of explaining the notion of an experience-independent warrant. Unlike his treatment of the concept of warrant in general, Kitcher seeks to give a detailed solution to this problem.

As Kitcher points out, we must take care in giving an account of experience-independence. It will not do to claim that a process is experience-independent just in case it could have produced a warranted belief given *any* background of experience: for this would entail that the process can produce a warranted belief given a background of *no* experience, and it is plausible that given such a background, no belief forming processes would exist at all, let alone produce warranted beliefs.⁴⁰

What is required is an account of experience-independence that will allow an appropriate role for experience: the account must respect the insight that we can know certain conceptual truths a priori, even though experience was required for us to acquire the concepts involved in those truths. In order to develop such an account, Kitcher introduces the idea of a *sufficient life*. The total sequence of experiences *S* has had up to time *t* is *S's life at t*. A life is *sufficient for S for p* just in case *S* could have had that life and gained sufficient understanding to be able to grasp *p*.⁴¹ We cannot, though, use this notion to give a definition of a priori knowledge as follows:

S knows *p* a priori just in case *S* knows *p* and for any life, sufficient for *S* for *p*, *S* could have that life and still know *p*.

⁴⁰ Kitcher 1983: 21.

⁴¹ Kitcher 1980: 5.

Naturalism and a priori knowledge

This proposed account has the consequence that where S knows p a posteriori, but *could* have known p a priori, then S knows p a priori. But this is unacceptable: it is possible to know a priori propositions a posteriori.⁴²

In looking for an account that works, Kitcher proposes to appeal to Kant's theory of mathematical knowledge through "pure intuition" and so give an account of the notion of experience-independence in general, and then to use Kant's suggestions, suitably formalised, in the definition of a priori warrant. In doing so, he does not seek to endorse Kant's theory: rather, he takes it to be illustrating the sorts of properties which belief forming processes must have if they are to be considered experience-independent. Kitcher writes:

On Kant's theory, processes of pure intuition are supposed to yield a priori mathematical knowledge. Let us focus on a simple geometrical example. We are supposed to gain a priori knoweldge [*sic*] of the elementary properties of triangles by using our grasp on [*sic*] the concept of triangle to construct a mental picture of a triangle and by inspecting this picture with the mind's eye. What are the characteristics of this kind of process which make Kant want to say that it produces knowledge which is independent of experience? I believe that Kant's account implies that three conditions must be met. The same type of process must be *available* independently of experience. It must produce *warranted* belief independently of experience. And it must produce *true* belief independently of experience.⁴³

Kitcher incorporates these three conditions in the following definition:

- (2) S knows p a priori iff S knows p , and S's belief in p was produced by a process which is an a priori warrant for it.
- (3) α is an a priori warrant for S's belief in p iff α is a process such that, given any life, sufficient for S for p
 - (a) some process of the same type could produce in S a belief in p ;
 - (b) if a process of the same type were to produce in S a belief in p then it would warrant S in believing in p ;

⁴² Cf. Kripke's discussion of gaining arithmetical knowledge a posteriori, via a calculator or computer (Kripke 1980: 35, discussed in chapter one of this thesis).

⁴³ Kitcher 1980: 23.

Naturalism and a priori knowledge

- (c) if a process of the same type were to produce in S a belief in p then p is true.⁴⁴

The analysis is intended to show that if S knows p a priori, then S could know p whatever background of experience she has had, provided that that experience is sufficient for S to acquire the concepts in p . Two observations in particular are pertinent.

First, regarding the individuation of *types* of processes, Kitcher says no more than that our intuitive principles of type-individuation must be respected.⁴⁵ An account which allowed the process of forming a belief in p after hearing a lecture, and the process of mentally running through a proof of p , to be tokens of the same type, would, he suggests, violate these intuitive principles. He writes that his analysis is meant to be read as a challenge to the apriorist.

If someone wishes to claim that a particular belief is an item of apriori knowledge then he must specify a segment of the causal ancestry of the belief, consisting of states and events internal to the believer, and type-identity conditions which conform to some principle (or set of principles) of classification which are standardly employed in our divisions of belief forming processes [...]. If he succeeds in doing this so that the requirements in (3) are met, his claim is sustained; if he cannot, then his claim is defeated.⁴⁶

Second, the *modality* incorporated into the definition by the inclusion of the word "could" in (3a) needs specifying. Without the modal operator, a claim that α is an a priori warrant would be defeated by exhibiting a scenario in which α never operates and so never produces any beliefs at all. But the required sense of possibility is not mere logical possibility, else (3a) would too weak. Kitcher suggests that the sense of "could" required should be restricted to possibilities in which the subject concerned has the "kinds of cognitive capacities distinctive of humans".⁴⁷ The idea behind (3a) is that in a world W (where W is a world in which S has a life sufficient to grasp p) a process α *could* produce belief in p in S just in case

⁴⁴ Kitcher 1980: 9-10; 1983: 24.

⁴⁵ Kitcher 1983: 24-6.

⁴⁶ *Ibid.*: 26.

⁴⁷ *Ibid.*

Naturalism and a priori knowledge

there is at least one possibility relative to W in which S has the kinds of cognitive capacities distinctive of humans, and α does produce a belief in p .

This reference to human-style cognitive capacities is to be built into conditions (3b) and (3c), which are to be modified accordingly.⁴⁸ Kitcher notes that this introduces vagueness as to which worlds are to be counted as relevant, but this vagueness mirrors our ignorance as to what distinctively human cognitive capacities are, and is neither avoidable nor objectionable.

It will be useful to illustrate Kitcher's definition with examples. We will consider two cases: one where Kitcher's account tests positive for apriority, and one where it does not.

For the first illustration, let the proposition p believed by S be that expressed by the sentence "I exist", as tokened by S . Intuitively, the process α that produces the belief in the actual world confers warrant on it: hence S knows p , and the first conjunct of condition (2) is met. Further, it seems clear that the process α - presumably a process of self-reflection - is such that a process of the same type can function and produce belief in p in any situation in which there is a background of experience sufficient for S to have grasped the concepts involved in p , and in which S has the kind of cognitive capacities distinctive of humans: hence (3a) is filled. What is more, there seems to be no way that a process of the same type as α could fail to warrant the belief in p , should it produce it; hence (3b) is filled. And finally, owing to its indexical character, p cannot be falsely tokened, so whenever α produces a belief in p , p is true. Hence (3c) is filled, and so S knows p a priori.

For the second illustration, let p be a complex mathematical truth. We can allow that in the actual world, S comes to believe p by proving it; hence, in the actual world S knows p . However: it seems to be possible that S could have a life sufficient to grasp p , and α be available for S to arrive at a belief in p , and yet *fail to warrant* S 's belief in p . Consider a situation in which S proves p , but then is misled by authoritative sources into thinking that her proof is incorrect (Kitcher calls this a *social challenge* to knowledge⁴⁹). Although her proof is correct, the authority of the

⁴⁸ *Ibid.*: 27.

⁴⁹ *Ibid.* see esp. chapter two.

Naturalism and a priori knowledge

sources is such that S is not warranted in believing that her proof is correct, and hence, though α is available, it does not confer warrant on the belief it produces. Hence, it seems, p is not known a priori.

This is a central and striking feature of Kitcher's definition of a priori knowledge: mathematical truths - at least those more complex mathematical truths susceptible to social defeat - are not known a priori.

Albert Casullo has objected that Kitcher's definition implausibly requires that a priori warrants be independent of their "standing conditions", where the standing condition for a process is the "complex neurophysiological state of a person which is nomologically necessary for a process to produce beliefs in that person."⁵⁰ He considers a counterexample to Kitcher's clause (3a) which is as follows.

Let α be the belief forming process consisting in the process of reflecting on the concepts of *triangle*, *line*, etc.; let a single brain state N be nomologically necessary for a subject to employ α , but *not* necessary for S to acquire the concepts of *triangle*, *line*, etc.; and let p be the belief that no two sides of a triangle are parallel. Now, suppose that in the actual world, S has a life sufficient to grasp p , and comes to know p by using process α . Suppose that in some world W^* , nomologically identical to the actual world down to the neural level, S gains all the concepts necessary to grasp p but lacks N . Hence, in W^* , S has a life sufficient to grasp p , but process α is not available to S . Hence, by (3a), α was not an a priori warranting process. But we have stipulated α and p in such a way that it just does seem that in the actual world, α warrants belief in p a priori. Hence, Casullo thinks, (3a) is implausible and must be modified.⁵¹

It seems likely, however, that Kitcher's remarks about the modality appealed to in his definition can handle this objection. In W^* , S is unable to reflect on the concepts of *line*, *triangle*, etc., to arrive at the belief that no two sides of a triangle are parallel. But it seems arguable that the ability to engage in this sort of geometrical reflection is a cognitive ability which is distinctively human. If this claim can be

⁵⁰ Casullo 1988: 202.

⁵¹ *Ibid.*: 203.

Naturalism and a priori knowledge

made out, it is open for Kitcher to claim that W^* does not represent a possibility in which S has the kinds of cognitive capacities distinctive of humans, and that it is therefore not relevant to whether α is an a priori process.

In response it could be argued that the mere unavailability to S in W^* of process α is not a sufficiently large change to S's cognitive capacities to warrant describing her as not distinctively human. But it is unclear whether we can make much headway with this problem by debating such examples: as we noted, we do not have a precise conception of what counts as distinctively human cognitive capacities, and without a precise conception we are unable to evaluate Casullo's objection to Kitcher. If Casullo's claim that Kitcher's clause (3a) must be modified is to be upheld, it will have to be supported by other arguments and it is not clear what these would be. I will therefore put this problem on one side.

Although Kitcher thinks that the concept of a priori knowledge is coherent, his account of the concept of experience-independence is so strong that the majority of things commonly thought to be a priori turn out not to be. In effect, then, Kitcher seems to vindicate the view that the naturalist must be hostile to a priori knowledge. Following the suggestions of authors who have commented on Kitcher, we will proceed by challenging the notion of experience-independence to which he appeals. To this end, in the next chapter, we will first locate Kitcher's ideas within a more general taxonomy of accounts of experience-independence; then we will examine the arguments for and against adopting Kitcher's account.

CHAPTER FIVE

Unrevisability

Kitcher's chief methodological insight is that the central problem for naturalised theories of a priori knowledge is that of giving a plausible account of what it is for a warrant to be experience-independent. This chapter embarks on the project of giving such an account, drawing from Kitcher in three further ways. First, Kitcher's own account of experience-independence will be used as a point of departure, and through raising problems with Kitcher's account, a more plausible theory will be developed. Second, Kitcher's notion of "warrant" - where a belief is warranted just in case it is produced by a belief forming process with appropriate characteristics - will be retained. Third, Kitcher's hope that an account of experience-independence can be given independently of considerations of the theory of knowledge in general will be maintained for as long as is practical. Ultimately, though, his attempt to segregate the issues in this way will prove untenable.

This chapter begins by sketching the options available to us in characterising the notion of an experience-independent warrant. Having done this, we consider whether the notion of *unrevisability* is suitable for characterising experience-independence. The first problem in determining whether it is suitable is to get a clear idea of what "unrevisable" means; this, it will transpire, requires an extended discussion of the concept of evidence. Having clarified what "unrevisable" means, and hence having clarified what is at stake in claiming that a priori knowledge is unrevisable, we turn to look at the arguments for and against that claim.

Experience-independence

We begin by giving a brief survey of the different ways in which a warranting process could be thought to be experience-independent. Current literature supplies three options in this regard.

Unrevisability

(i) The first option analyses experience-independence in terms of the *inputs* to the belief forming process which is the warrant. On this account, a warranting process is independent of experience if it has *nonexperiential* inputs: accordingly, we shall say that on this option, experience-independence is analysed in terms of *nonexperientiality*. It is not immediately obvious what the notion of "nonexperiential input" will involve; however, we will not discuss this problem until chapter nine.

(ii) The second option analyses experience-independence in terms of the *strength of the warrant* conferred on a belief produced by the belief forming process. On this account, a warranting process is independent of experience if the beliefs it produces are *unrevisable*.¹ Roughly speaking, a belief is unrevisable just in case it would never be rational to give it up. As we will see, this rough definition masks some serious problems.

(iii) The third option analyses experience-independence in terms of the *reliability* of the warrant. On this account, a warrant is independent of experience if it is *infallible*, that is, if it never attaches to a false belief. A priori warrant is infallible just in case no false beliefs are ever a priori warranted. In what follows, *infallibilism* will be the view that a priori warranting processes are infallible.

The first of these three accounts is the *minimal notion* of experience-independence: intuitively, a belief forming process that needs experiential input in order to produce warranted beliefs is certainly not experience-independent. The minimal notion of experience-independence can be augmented by one or both of the remaining options to produce a stronger account. Kitcher, as we will see, builds all three into his account of experience-independence; later authors have sought to defend an account of experience-independence defined solely in terms of the minimal notion.

¹ This approach is adopted by Kitcher 1980 and Putnam 1975, 1983, and is endorsed implicitly by Quine.

Unrevisability

As noted, this chapter is centrally concerned with the second notion of experience-independence, that is, unrevisability. Whether infallibility is a suitable notion for defining experience-independence will be treated of in the following chapter. The discussion of infallibility will expose a serious problem for the naturalistic theory of knowledge in general, which is addressed in chapters seven and eight. Having addressed this problem, we return, in chapter nine, to the minimal notion of experience-independence.

The concept of unrevisability

We define "unrevisable" for a belief as follows:

- (1) in w , S 's belief in p is *unrevisable* iff there is no situation accessible from w in which it is rational for S to give up her belief in p .²

Following Casullo³, we will refer to the thesis that a priori knowledge is unrevisable knowledge as "(UT)":

- (UT) If S 's belief in p is warranted a priori then S 's belief in p is unrevisable.

This definition of "unrevisable" draws from an account offered by Putnam. Putnam writes: "...an unrevisable statement is one that one would never be rational to give up".⁴ But this definition stands in need of clarification. Our attention should be drawn to the accessibility relation appealed to in (1). If this relation is set too weak, virtually no beliefs will count as unrevisable. Even a proponent of (UT) will want to admit that there are truths now known a priori that it was rational not to believe in at a time *before* the beliefs became warranted: it may well not be rational for S , as a six-year-old, to believe in Pythagoras' Theorem, even though S may come

² It is common in characterising unrevisability to draw a distinction between prudential and epistemic rationality; it is the latter sense of "rationality" which is appealed to here. It may be prudentially rational to give up a belief when faced with a serious threat, but this would not bear on the epistemic status of the belief.

³ Casullo 1988.

⁴ Putnam 1983: 98.

Unrevisability

later in life to know that theorem a priori. If S somehow found herself back in her 6-year-old cognitive state, augmented only by a belief in Pythagoras' Theorem, it might well be rational for S to give up that belief. If unrestricted, the accessibility relation will generate an unrevisability thesis that requires that a priori truths are *always* rationally believed; but it is unfair to the proponent of (UT) to lumber them with such an implausible thesis.

We need a principled way of strengthening the relation, so as to generate an unrevisability thesis which has at least initial plausibility. A temporal restriction - such that only situations which occur subsequent to the formation of the putatively unrevisable belief are accessible - will not help. On such a restriction, (UT) becomes the thesis that a priori knowledge, once acquired, can never be lost; but this too seems to be an unreasonably restrictive thesis: intuitively, it seems plausible to allow that at some point in time S can prove Pythagoras' Theorem and so come to know it a priori, and then at a later time forget how she proved it and so cease to know it. We can forget things: and if S forgets enough it could well become rational for her to give up her belief in Pythagoras' Theorem.

A more promising approach would begin by recalling that the concept of unrevisability was appealed to as a way of explaining the concept of experience-independence; with this in mind, we might consider restricting the accessibility relation by appeal to conditions on S's evidence in the situations to be deemed accessible. The intuitive idea that we are trying to capture is that an unrevisable belief is one which, once acquired, cannot be dislodged by the *accumulation of additional* evidence. This suggests that the accessible situations should be those wherein **S's evidence for p is undiminished**. This generates an unrevisability thesis according to which an item of a priori knowledge is supported by the evidence in such a way that no further evidence can undermine it, leaving open the possibility that it can be undermined by loss of evidence. That is, we have the following definition:

- (2) in w , S's belief in p is *unrevisable* iff there is no situation accessible from w in which S's evidence for p is undiminished and in which it is rational for S to give up her belief in p .

Unrevisability

Clearly, this is the sort of unrevisability thesis that we are interested in. However, the suggestion carries with it a number of problems.

First, it calls for a way of reidentifying the specific evidence for a belief in situations where the totality of evidence is different. This will be problematic, for what a proposition p is evidence for depends on the context. Is the proposition *Beth is going out the door at the back of the hall* evidence that she is going to get the ball? If we have just decided to play football, and I know that the ball is kept in the cupboard at the back of the hall, then it may be. If I am holding the ball in my hand, then it is not. If the present account of accessibility is to be adopted, therefore, some care will have to be taken to give an account of the reidentification of evidence in different situations. Since I doubt that the concept of unrevisability can be of much help in defining experience-independence, I do not propose to attempt such an account.

Second, the plausibility of the definition of "unrevisable" given in (2) is sensitive to what we take evidence to be. On an historically influential view of evidence, evidence is a form of experience; a common variation on this theme has it that evidence proper is not misidentifiable. To deny this latter claim seems to generate a regress. For, suppose that my evidence was misidentifiable; then, to be sure that a given item was part of my evidence, I would need more evidence; but to be sure that another given item was part of this new evidence, I would need more evidence, etc.⁵ This regress needs to be stopped "with a bullet", by the postulation of a level of evidence which has a privileged epistemic status: we cannot be wrong about it.

This conception of evidence - as an epistemically significant experience which "is what it appears to be" - has consistently eluded attempts to clarify it, and it would be unwise to yoke our notion of unrevisability to a potentially incoherent notion of evidence.⁶ A more promising account, developed by Williamson, identifies S 's evidence with the set of propositions S knows.⁷ Williamson calls this

⁵ See Williamson 1997: 721.

⁶ See in particular Sellars 1963, Williamson 1996a.

⁷ Williamson 1997.

Unrevisability

thesis $E=K$. The thesis that all and only knowledge is evidence is equivalent to the conjunction of three claims: (i) all evidence is propositional, (ii) all propositional evidence is knowledge, (iii) all knowledge is evidence.⁸ Williamson argues for (i) and (ii) by identifying the central functions of the concept of evidence and seeing what serves them.

Arguments for (i)

The concept of evidence is central to inference to the best explanation, probabilistic reasoning and the evidential exclusion of hypotheses. In inference to the best explanation we choose between hypotheses on the basis of which of them best explains our evidence⁹; but what hypotheses explain is propositional: hence evidence is propositional.

The conditional probabilities of hypotheses h and h^* on evidence e can be compared by calculating the inverse probabilities of e on h and h^* ; this involves assigning probabilities to evidence; but where, as in the present case, the probability in question is a measure of the evidential status of a belief, what is assigned a probability is a proposition. Hence, again, evidence is propositional.

Evidence e rules out an hypothesis h iff $e \vdash \neg h$. But the premises of an argument are propositions, hence, evidence is propositional.

Arguments for (ii)

Again, Williamson argues via considerations of inference to the best explanation, probabilistic reasoning and the evidential exclusion of hypotheses.

When we prefer h to h^* because h explains e better than h^* does, we are standardly assuming e to be known; if we do not know e , why should h 's capacity to explain e confirm h for us? It is likewise hard to see why the probability of h on e should regulate our degree of belief in h unless we know e . Again, an incompatibility between h and e does not rule out h unless e is known.¹⁰

⁸ *Ibid.*: 724.

⁹ Lipton 1991.

¹⁰ Williamson 1997: 731.

Unrevisability

The possibility of performing a "Russellian Retreat" at this point makes it necessary to consider this argument more closely. A "Russellian Retreat" is a retreat, typically under sceptical pressure, from claims about knowledge to claims about justified belief.¹¹ In the present context, the possibility seems open to argue that not all propositional evidence is knowledge, through attempting to demonstrate that the central functions of evidence can be subserved by something other than knowledge. The obvious candidate here is justified true belief; in response, Williamson argues that justified true belief cannot discharge the functions of evidence in the way that knowledge can.

Suppose that balls are drawn from a bag, with replacement. [A]ssume that someone else has already made the draws; I watch them on film. The following situation can arise. I have seen draws 1 to N ; each was red (produced a red ball). I have not yet seen draw $N+1$. I reason probabilistically, and form a justified true belief that draw $N+1$ was red too. My belief is in fact true. But I do not know that draw $N+1$ was red. Consider two false hypotheses:

h : Draws 1 to N were red; draw $N+1$ was black.

h^* : Draw 1 was black; draws 2 to $N+1$ were red.

It is natural to say that h is consistent with my evidence and that h^* is not. In particular, it is consistent with my evidence that draw $N+1$ was black; it is not consistent with my evidence that draw 1 was black. Thus my evidence does not include the proposition that draw $N+1$ was red. Why not? After all, I have a justified true belief that it was red. The obvious answer is that I do not *know* that draw $N+1$ was red; the unsatisfied necessary condition for evidence is knowledge.¹²

Hence, it seems implausible that something less than knowledge can subserve the functions required of evidence.

Arguments for (iii)

Since, intuitively, we want as much evidence as possible, the burden of proof here is on those who would deny that all knowledge is evidence.¹³

¹¹ Wright 1991.

¹² Williamson 1997: 731.

¹³ *Ibid.*: 734.

Unrevisability

However, this claim generates a problem for the present conception of unrevisability. Given component (iii) of $E=K$, the claim that all knowledge is evidence, it follows that every item of knowledge is evidence for itself.¹⁴ Hence, the proponent of $E=K$ is committed to the claim that it is right to respond to the question "what is the evidence for e ?" by citing e . It would certainly be strange to respond to the question "what is the evidence for e ?" in this way; however, to establish a counterexample to $E=K$ we will need to show more than that it is strange that e is evidence for e : we will need to show that it is false. But as Grice has pointed out, that an utterance is strange is not sufficient for it to be false: it may simply be inappropriate in the conversational context.¹⁵ Appealing to this distinction, Williamson maintains that the strangeness of answering " e " to the question above shows only that the response would be inappropriate, not that it is false. e is indeed evidence for e .

However, the claim that each item of knowledge is evidence for itself, in conjunction with the current definition of "unrevisable", entails that *all knowledge is unrevisable*. We can show this by arguing as follows, for an arbitrary p .

- | | | |
|-------|---|---|
| (i) | S knows p . | <i>Assumption</i> |
| (ii) | S's belief in p is unrevisable iff there is no situation in which S's evidence for p is undiminished and in which it is rational for S to give up her belief in p . | <i>Abbreviated definition of unrevisability</i> |
| (iii) | If S knows p then it is not rational for S to give up her belief in p . | <i>Conceptual truth</i> |
| (iv) | Every item of knowledge is evidence for itself | <i>From $E=K$</i> |
| (v) | p is part of S's evidence for p . | <i>From (i) and (iv)</i> |

¹⁴ *Ibid.*: 735.

¹⁵ Grice 1989. We should not think that this is a distinction of interest only to philosophers, as the following quotation illustrates:

"Who's the girl you want it for?"

"She's about the same age as us," Christopher said and, since Oneir was looking at him for a further explanation and he was fairly sure Oneir was not going to believe in someone called the Goddess, he added, "I've got this cousin called Caroline." This was quite true. Mama had once shown him a studio photo of his cousin, all lace and curls. Oneir was not to know that this had nothing whatsoever to do with the sentence that had gone before.

From *The Lives of Christopher Chant* by Diana Wynne Jones.

Unrevisability

- (vi) In every situation in which S's evidence for p is undiminished, p is part of S's evidence. *Obvious consequence of (v)*
- (vii) In every situation in which S's evidence for p is undiminished, S knows p . *E=K, applied to (vi)*
- (viii) There is no situation in which S's evidence for p is undiminished and in which it is rational for S to give up her belief in p . *From (iii) and (vii)*
- (ix) S's belief in p is unrevisable *From (ii) and (vii)*

But p was arbitrary: therefore all knowledge is unrevisable.

The claim in line (iii) might be contested: but if it is false then its contrapositive is also false. The contrapositive is the conditional: *if it is rational for S to give up her belief in p then S does not know p* ; but this seems supported by the commonplace observation that "knowledge excludes doubt". Therefore it seems (iii) must be sustained, and, if we grant $E=K$, we can only respond to this argument by placing a further restriction on the accessibility relation in the definition of unrevisability. One way of imposing this further restriction would be to appeal to Williamson's notion of the *independent* evidence for a proposition. The independent evidence for p is evidence that one could have for p without knowing p .¹⁶ One could infer from the independent evidence for p to p , but this inference is clearly such that the addition of further evidence can undermine it.¹⁷ Since a revisable item of knowledge is precisely one that can be undermined by additional evidence, this seems to capture what we need of the concept of unrevisability without making all knowledge unrevisable: an unrevisable item of knowledge is one that no additional evidence could undermine.

The completed definition of unrevisability is therefore as follows:

¹⁶ Williamson 1997: 735.

¹⁷ The independent evidence for p is defined as the evidence that one *may* have for p without knowing p ; if one's evidence for p included p itself, then one *could not* fail to know p on the basis of one's evidence for it. Hence it is not unreasonable to assume that one may often, though not always, know p on the basis of the independent evidence for it.

Unrevisability

- (3) In w , S 's belief in p *unrevisable* iff there is no situation accessible from w in which S 's independent evidence for p is undiminished and in which it is rational for S to give up her belief in p .

The definition still suffers from the earlier problem of reidentifying the evidence for p in situations where the total evidence is different; but as before, since I do not think that an appeal to unrevisability can help us give an account of experience-independence, I propose to let this pass.

Though it is valuable to have a clear account of the concept of unrevisability, the rough formulation given in (1) will in fact be all we need for the following discussion. Therefore, I will revert to that formulation in what follows.

The unrevisability thesis

We now turn to examine arguments for and against the unrevisability thesis.

Referring back to Kitcher's definition of a priori knowledge given in the previous chapter, it is easy to see that Kitcher adopts the notion of unrevisability as a criterion of experience-independence. For, assume that an a priori warranting process α produces in S a revisable belief in p ; then, α has produced a belief such that there is a possible evidential situation e in which it is rational for S to give up the belief in p ; so, in e , S is not warranted in believing in p ; hence, in e , a process of the same type as α could produce in S a belief p and not warrant S in believing in p . So by Kitcher's clause (3b), α was not an a priori warranting process. Thus, on Kitcher's definition, all beliefs produced by a priori warranting processes are unrevisable.

So, clearly, Kitcher endorses (UT). The question is: what justification does he offer for this?

Kitcher's attempt to motivate the unrevisability thesis

(UT) is not a trivial thesis. For a start, it is quite plausible that knowledge in general is revisable. That is, the following proposition is false:

Unrevisability

If S knows p , then S's belief in p is unrevisable.¹⁸

So (UT) does not fall straight out of considerations of knowledge in general. Nor does (UT) follow from uncontroversial propositions regarding knowledge in general and a priori warrant, as Casullo illustrates by considering the following claims:

- (i) A priori warrant is experience-independent warrant.
- (ii) The existence of a priori knowledge entails that there is experience-independent warrant sufficient for knowledge.
- (iii) The general concept of knowledge does not require that possession of warrant sufficient for knowledge entail unrevisability.
- (iv) It is not the case that if S is warranted a priori in believing p then S's belief in p is unrevisable.

This is not meant as an argument, but as an illustration. These propositions are consistent, and (i), (ii), and (iii) are plausible observations about knowledge and a priori warrant. Yet (iv) is the negation of (UT). So if (UT) is to be established, special argumentation will have to be adduced.¹⁹

Why, then, does Kitcher think that (UT) is true? As we saw in the previous chapter, he takes it that his definition of a priori knowledge captures the features that had made authors such as Kant want to claim that a priori knowledge is experience-independent knowledge. But - even granting Kitcher's interpretation of Kant's epistemology - the reference to Kant is so far no more than an argument from authority: we have been given no reason to accept Kant's views on experience-independence. However, Kitcher has an independent argument for defining experience-independence in terms of unrevisability, which we will now consider.

As noted in the previous chapter, Kitcher's intention in his 1983 work is to argue against apriorism about mathematical knowledge by showing that most

¹⁸ Casullo 1988: 188.

¹⁹ *Ibid.*: 198-190. I have substituted "warrant" for Casullo's use of "justification", to avoid potential confusions regarding issues that will be resolved in the following chapter.

Unrevisability

mathematical knowledge is revisable, and so fails the criteria for a priori knowledge set out in his definition. Having made his case for this, Kitcher considers the charge that his definition of experience-independence imposes unreasonable requirements on a priori knowledge, and that a reasonable definition of a priori knowledge can be framed that concedes its revisability. He then writes:

To make this concession is to abandon the fundamental idea that a priori knowledge is knowledge which is independent of experience. The apriorist would be saying that one can know a priori that p in a particular way, even though, given appropriate experiences, one would not be able to know that p in the same way. But if alternative experiences could undermine one's knowledge then there are features of one's current experience which are relevant to the knowledge, namely those features whose *absence* would change the current experience into the subversive experience. The idea of the support lent by kindly experience is the obverse of the idea of the defeat brought by uncooperative experience. To reject condition (3b)... would be to strip apriorism of its distinctive claim.²⁰

Kitcher's claim is striking. In effect, he is arguing that the minimal account of experience-independence, given in terms of nonexperientiality, cannot be distinguished from the apparently stronger account in terms of unrevisability! To see this, note first that it is generally held that beliefs which are produced by *experiential* processes are revisable. Hence, the following conditional seems to be true:

$$\forall p (p \text{ was produced by an experiential process} \rightarrow p \text{ is revisable})$$

Kitcher argues for the converse of this as follows. Assume that S's warranted belief in p is revisable. Then there is some feature of experience e such that if e obtains, S's warrant is defeated. Hence, if S's belief in p is warranted, then S's experience lacks e . If e is a feature of experience, then the lack of e is also a feature of experience. So, that S's experience lacks e must be considered causally relevant to the production of S's belief in p . Hence, there is a feature of experience which must be considered

²⁰ Kitcher 1983: 88-9.

Unrevisability

causally relevant to the production of S's belief in p . Hence S's belief in p is produced by an experiential process. But p was arbitrary. Hence we seem to have established:

$$\forall p (p \text{ is revisable} \rightarrow p \text{ was produced by an experiential process}).$$

These two conditionals combine to generate the following claim:

$$\forall p (p \text{ is revisable} \leftrightarrow p \text{ was produced by an experiential process}).$$

Whence, clearly, all and only unrevisable beliefs are produced by nonexperiential processes. Hence, Kitcher concludes that any attempt to capture the notion of independence from experience must build in unrevisability.

Summerfield, however, has argued that Kitcher's argument conflates different notions of "dependence". Summerfield maintains that though Kitcher may be right to say that the lack of e must be considered a feature of experience, he is wrong to claim that the lack of e must be considered causally relevant to the production of S's belief. Her point turns on recognising that there is an intuitive difference between the cause of an event, the background conditions against which the cause was efficacious, and further conditions which, had they obtained, would have blocked the ability of the cause to produce the effect. Although it is not a straightforward matter to make this difference precise, it is likely that any satisfactory account of causation will have to respect it.²¹ In general, where event²² c causes event e at time t , and for a condition x which does not occur but which could, by occurring at t , have prevented c from causing e : the fact that there is a counterfactual situation in which x does occur at t and so blocks c causing e does not make x part of the cause of e in the actual world.

²¹ Summerfield 1991: 43.

²² I assume that the casual relata are events; a defence of this assumption would take us beyond the remit of this thesis, nor do we have to establish this for the present point to retain its force.

Unrevisability

An example might make this clearer. A bin fire in a building causes a conflagration which razes that building to the ground; had the building been flooded the bin fire would not have caused the conflagration; but the non-obtaining of the flooding does not thereby become part of the cause of the conflagration. It remains simply a condition which, had it obtained, *would have* prevented the conflagration. The intuitive distinction between cause and background condition gives us a reason to deny causal status to the non-floodedness of the building.

Analogously, the non-occurrence at the moment of the formation of a belief in *p*, of experiences *e* which would prevent the formation of that belief, cannot be taken as showing that *p* is produced by an experiential process. Summerfield's observations give us a principled way of maintaining the distinction between unrevisability and nonexperientiality; and so blocking Kitcher's argument for (UT).²³

Casullo's argument against the unrevisability thesis

We now move on to arguments that (UT) is not just unsupported, but actually false. Objections to (UT) have been developed by Casullo.

Casullo notes first that it is natural to think that there is a distinction between experiential evidence and nonexperiential evidence; further, it is natural to think that the fact that a belief is revisable in the face of *nonexperiential* evidence need not count against that belief's being warranted a priori. Casullo sees this observation as pointing to a distinction between "strong" and "weak" unrevisability theses:

(SUT) If S's belief in *p* is warranted a priori then S's belief in *p* is unrevisable in the light of *any* future evidence.

(WUT) If S's belief in *p* is warranted a priori then S's belief in *p* is unrevisable in the light of any future *experiential* evidence.²⁴

²³ See also Casullo's attribution to Kitcher of "symmetry thesis" about confirming and defeating evidence, and his counterexamples to such a thesis: Casullo 1988: 197-9.

²⁴ Casullo 1988: 190.

Casullo thinks that the strong version is particularly implausible. He writes:

Clearly, (WUT) is more plausible than (SUT). For suppose that S's belief that *p* is justified on the basis of nonexperiential evidence and it is acknowledged that *p* might be rationally revised in light of further *nonexperiential* evidence. In such a case it does not appear plausible to maintain that S's justification is *not* a priori. (WUT) is more promising since one can argue that if S's belief that *p* is revised in light of *experiential* evidence then that belief is not independent of experience in the requisite sense.²⁵

Accordingly, Casullo quickly rejects (SUT) and directs the bulk of his arguments against (WUT). However, an argument due to Edidin suggests strongly that the distinction between (SUT) and (WUT) is untenable. Edidin's argument can be read as showing that a belief is strongly unrevisable if it is unrevisable at all. The argument, recast to fit present terminology, is as follows.²⁶

Let S have an a priori warranted belief in *p*; suppose S's belief in *p* is revisable in the face of nonexperiential evidence. Then, there is no reason why S could not acquire evidence which warrants her in believing that *p* is revisable by nonexperiential evidence and which warrants her in believing that such evidence could be obtained by persons other than herself. Then let T be an individual whom S believes to be extremely authoritative regarding beliefs susceptible to warrant by nonexperiential evidence, and let S know that T is familiar with S's warrant for believing *p*. T then says that she possesses nonexperiential evidence that undermines S's warrant for believing *p*, though she does not say what that evidence is. Given S's respect for T, S would be irrational to persist in believing *p*; hence this is a situation in which it is rational for S to stop believing *p*. And the evidence which leads her to do so is testimonial, and therefore experiential.²⁷ Hence, if S's belief in *p* is revisable in the face of nonexperiential evidence, S's belief in *p* is also revisable in the face of experiential evidence, and therefore not unrevisable at all.

²⁵ *Ibid.*

²⁶ Edidin 1984: 190.

²⁷ This assumes that testimonial evidence is experiential evidence. This assumption is not uniformly accepted (Cf.: Burge 1993); nevertheless, it has a great deal of plausibility and I will not question it in this thesis.

Unrevisability

If a warranted belief is not unrevisable in the face of both experiential and nonexperiential evidence, it is not unrevisable at all. Contraposing: if a belief is unrevisable it is unrevisable in the face of both experiential and nonexperiential evidence; that is, it is strongly unrevisable. So, if Edidin's argument is sound, (WUT) does not represent a stable unrevisability thesis, and we should reject the suggestion that a priori knowledge can be characterised by (WUT). (SUT) is the only tenable unrevisability thesis: in effect, the thesis (UT) *just is* the thesis (SUT).

We should note, though, that there does indeed seem to be an intuitive distinction between experiential and nonexperiential evidence. Nonexperiential evidence seems to be just that sort of evidence which can produce a priori knowledge.²⁸ The argument above does not show that there is no distinction between these two sorts of evidence. What it shows is that a belief must be unrevisable in the face of both sorts of evidence if it is unrevisable at all.

The crucial parts of Casullo's argument, then, are his considerations against (SUT). Casullo proposes the following case as a counterexample to (SUT).

- A Mary is able reliably to discriminate between valid and invalid inferences on the basis of reflective thought. Mary wonders whether " $p \rightarrow q$ " entails " $\neg p \rightarrow \neg q$ " and at t concludes that it does. Later, a counterexample occurs to her which leads her to reject the belief she held at t , and to believe instead that " $p \rightarrow q$ " entails " $\neg q \rightarrow \neg p$ ".²⁹

(Note that it would not do to respond to this case by saying that Mary's erroneous belief at t shows that she is in fact not reliable: intuitively, warrant can be conferred

²⁸ This distinction is not universally accepted: see for example Butchvarov 1970. Whether we allow a distinction between a priori and non-a priori evidence depends on one's theory of evidence. For Quine, for example, evidence is the stimulation of sensory receptors (Quine 1969: 75); on an account such as this, it would make little sense to speak of a priori evidence, since a priori connotes "nonsensory". On Williamson's 1997 account of evidence, one's evidence consists of all and only the propositions that one knows; Williamson's account therefore allows for a distinction between a priori and non-a priori evidence: one's a priori evidence is all and only one's a priori knowledge.

²⁹ Casullo 1988: 191.

Unrevisability

by processes which are reliable but not infallible.³⁰) Casullo makes five claims about case A; three are banal, *viz.*:

- (a) Mary's belief at *t* is based on a nonexperiential process α which is reliable but not infallible;
- (b) a process of the same type as α leads Mary to conclude that the belief she held at *t* is mistaken, and leads her to a correct belief;
- (c) Mary's conclusions as stated in (b) are warranted beliefs.

The final two are controversial:

- (d) Mary's belief at *t* was a warranted belief;
- (e) Mary's belief at *t* was warranted a priori despite later being revised.

In favour of (d) Casullo cites the following case, which seems to be analogous in all relevant respects to case A:

- B Mary sees a sheet of paper on the table and on that basis forms the belief that it is square. A second closer examination reveals that two of the sides of the sheet of paper are slightly longer than the other two. On that basis, Mary rejects her former belief about the shape of the paper and comes to believe that it is rectangular.³¹

Casullo argues as follows.

Since the circumstances under which Mary perceived the page were normal and Mary is a reliable discriminator of shapes, her initial belief is warranted. The fact that our discriminatory powers sometimes fail us does not entail that beliefs based on shape-perception are not warranted. Furthermore, if such beliefs typically are warranted, we do not single out particular cases as unwarranted merely because they are false. Some other relevant difference must be cited such that the perceiver was impaired or the environment was gerrymandered.³²

³⁰ This intuition is challenged in the following chapter.

³¹ Casullo 1988: 191.

³² *Ibid.* I have again substituted "warranted" for Casullo's "justified", to retain continuity with the text. The question of what concept of positive epistemic status these terms are placeholders for will shortly become relevant.

Unrevisability

From this it seems to follow that we must assent to (d). The proponent of (UT) must therefore deny (e): that is, it must be claimed that although Mary's belief at *t* is warranted, it is not warranted a priori, simply because it was revised. But, Casullo claims, this is very implausible. The process which leads Mary to revise her belief is, like the process which leads her to form it, nonexperiential. If we then deny that Mary's belief at *t* is warranted, we are ignoring the factor that is the chief determinant of apriority.

To this it might be objected that Casullo simply begs the question against the proponent of (UT). For the proponent of (UT), nonexperiential production is precisely *not* the sole determinant of whether a belief is a priori: it must also be unrevisable. But this charge of question begging only has force if we have independent reason to believe that (UT) may be true; and our recent consideration of Summerfield's arguments suggest that no such reason is forthcoming. Therefore the charge of question begging in this context can be discarded.

However, as Casullo points out, case A could be objected to on the grounds that it assumes that a priori warrant is fallible: that is, it assumes that there can be a priori warrant for false beliefs. Now, the proponent of (UT) must certainly claim that a priori warrant is infallible. To see this, let us assume both the thesis (UT), and that a priori warrant is fallible: and now suppose *S* has a false but a priori warranted belief in *p*. It follows from (UT) that *S* can *never* revise her belief in *p*, that she can *never* discover that her belief is false. But this is not a condition that we would think governs the sorts of beliefs we normally say are known a priori. Suppose *S* is an expert mathematician and convinces herself she has proven some very complex mathematical proposition *p*, which is in fact false; if her pseudo-proof is a priori warrant conferring, and a priori warranted beliefs are unrevisable, then *S*'s mistake will be uncorrectable! But since *S* was able to arrive at the belief in the first place, we are inclined to suppose that she has the ability to correct her belief if it is wrong. So the unrevisability of a priori warranted beliefs and the fallibility of a priori warrant are two doctrines it is implausible to hold simultaneously. Thus, someone who accepts that a priori knowledge is unrevisable must suppose that a priori warrant is infallible; so Casullo's assumption that there can be a priori

Unrevisability

warranted false beliefs threatens to beg the question against the proponent of (UT) in another way. Further, this charge of question-begging is dialectically more serious since we have not yet seen any arguments for or against infallibilism.

Casullo tries to sidestep this problem by suggesting that other cases can be set up which make the same point as case A, but which do not assume that a priori warrant is fallible. In an attempt to show this, Casullo offers the following case, meant to make the same point as case A, without assuming that a priori warrant is fallible.

- C Charlie believes p entails q on the basis of a valid proof P_1 . Since the proof is a result of reflective thought, Charlie's belief is warranted nonexperientially. But now suppose that (a) there exists a pseudo-proof P_2 from p to $\neg q$; and (b) that if this pseudo-proof were brought to Charlie's attention he would not be able to detect any flaws in it or to discount it in any other fashion.³³

Casullo intends to argue from this case to substantially the same conclusions as before. He asks us to imagine that P_2 is never brought to Charlie's attention: hence Charlie's belief that p entails q remains warranted despite the fact that it *would* be revised if P_2 did come to Charlie's attention. Charlie's belief is revisable, and it seems on the surface that that no false beliefs need to be warranted for this example to go through. Hence, Casullo concludes, there can be a priori warranted beliefs which can be revised.

But on reflection we might well be concerned that Casullo has failed to remove the assumption that a priori warrant is fallible. A priori warrant is infallible, recall, just in case a priori warrant never attaches to a false belief. But the process of following the pseudo-proof P_2 must be capable of conferring some positive epistemic status on Charlie's false belief that p entails $\neg q$ if it is to be capable of forcing Charlie to revise his true belief that p entails q . Now, P_2 is a mere pseudo-proof, therefore the process of following it is probably incapable of conferring any real positive epistemic status on beliefs. But it must be capable of conferring some apparent positive status on Charlie's belief that p entails $\neg q$: in some sense, Charlie is justified in believing p entails $\neg q$ after following the pseudo-proof, since as far as he

³³ *Ibid.*: 192.

Unrevisability

is concerned P_2 is a legitimate proof. So the process of following the pseudo-proof seems capable of conferring some form of positive epistemic status on a false belief. The question is whether this status is that of being warranted.

Our difficulty here is that we have not yet developed the conceptual tools needed for distinguishing real from apparent positive epistemic status³⁴, and that we need those tools in order to evaluate whether Casullo is successful in removing the assumption that a priori warrant is fallible. I propose, therefore, to set aside this line of argument and turn to alternative arguments against (UT) which do not make assumptions about the fallibility of a priori warranting processes.

The argument from the revisability of knowledge in general

Perhaps the strongest reason to believe that a priori knowledge is revisable derives from considerations from the theory of knowledge in general, and so is obscured by Kitcher's proposal that the analysis of "experience-independence" proceed independently of the wider theory of knowledge. The relevant consideration is that *all knowledge is revisable*. Given that a priori knowledge is knowledge, this suggests a very quick argument for the revisability of a priori knowledge: all knowledge is revisable, a priori knowledge is knowledge; therefore, a priori knowledge is revisable. We can call this "the argument from general revisability".

The proponent of (UT) might object that the claim that all knowledge is revisable simply begs the question against (UT). Again, this point would have some force if we had seen a plausible argument for (UT); given that we have not, the charge of question-begging is weak.

However, whether the argument from general revisability is ultimately convincing depends on the way that first premise is established. The premise is often called "Peircean fallibilism": this is the thesis that we should hold all our knowledge in as modest a way as possible, never considering any piece of knowledge that we have as immune to revision.³⁵ It is not clear, though, that this

³⁴ The distinction between warrant and the intuitive notion of justification is discussed in the following chapter.

³⁵ Cohen 1988, Casullo 1988.

Unrevisability

formulation really motivates the view that all knowledge is revisable, as opposed to the view that all knowledge should be *considered* revisable: it is consistent with our *holding* that a belief is revisable, that it is in fact unrevisable. The premise is also associated with Quine's holism, which we have already discussed at length. However, we can eschew the details of Peircean interpretation or further exegesis of Quine, for the premise is supported directly by the naturalistic framework we have adopted.

According to the naturalist framework, knowledge is produced by belief forming processes with a certain characteristic. Pending further discussion in the next chapter, let us take it that this property is reliability; a *reliabilist* is someone who endorses this view of the nature of the desired characteristic. To see how the reliabilist framework supports the view that all knowledge is revisable, we must first note a further distinction between types of revisability. There are two ways in which experience might lead us to revise beliefs that we take ourselves to know. Where warranting process α produces in S a belief p : we call *direct* defeaters, experiences that provide S with reason to believe that p is false; and we call *indirect* defeaters, experiences that provide S with reason to believe that α is not an appropriate (i.e.: reliable) belief forming process.³⁶ This leads us to recognise a distinction between two types of revisability:

S's belief in p is *directly revisable* iff there is an evidential situation in which S's evidence gives S reason to believe $\neg p$.

S's belief in p is *indirectly revisable* iff there is an evidential situation in which S's evidence gives S reason to believe that S's belief in p is not warranted.

If a belief is directly revisable then it is indirectly revisable but not *vice versa*; one can have evidence that one's belief is incorrectly formed without having evidence that one's belief is false.

³⁶ This distinction is due to Casullo 1988: 204.

Unrevisability

Now, we are calling cases wherein an authoritative source gives us (possibly misleading) reason to think that we do not know the things we took ourselves to know, cases of *social challenge* to knowledge.³⁷ The belief forming processes apt for producing knowledge are, on the naturalistic account, realised by neurophysiological processes, and we rationally have to respect the pronouncements of neurophysiologists on matters of neurophysiology. Hence, there are individuals who are in a more authoritative position with respect to our brain processes than we are, and we are rationally required to respect what they say on these matters. It follows that the vast majority of our knowledge is vulnerable to social challenges, for neurophysiologists can always give us (possibly misleading) reason to think that the processes that produce our knowledge are not reliable. And this observation applies equally to a priori knowledge. As Casullo writes:

...it can be plausibly argued that if one has constructed a valid proof for a particular theorem then the warrant conferred on the theorem by the process of constructing the proof cannot be defeated by experiences such as the testimony of authorities or the results of a computer programme. If one has a proof in hand then one is warranted in being suspect about the sincerity or competence of the alleged authorities and computer programmers. But when we turn to *indirect* defeaters the situation changes radically. First of all, it is generally granted by proponents of reliabilism that the warrant which a reliable process confers on S's belief that *p* is defeated if S has reason to believe that the process is *not* a reliable one. Secondly, the reliability of any cognitive process is a matter which is open to empirical investigation. Hence, there is some set of *possible* experiences which would justify us in believing that it is unreliable. Here it is crucial to recognise that even if a belief forming process is in fact reliable, it does not follow that the *available* evidence will warrant us in believing that the process is reliable.³⁸

Hence, it seems that even a priori knowledge is revisable, though Casullo admits that it is only indirectly revisable. However: indirect revisability is sufficient for revisability.

One obvious way of challenging this argument is to observe that the claim that all knowledge is revisable is just false. There are some beliefs which do seem to

³⁷ The terminology is Kitcher's (Kitcher 1983) and was introduced in the previous chapter.

³⁸ Casullo 1988: 205.

Unrevisability

be totally unrevisable: for example, the Cartesian *cogito*; or the belief that not every sentence is both true and false at the same time.³⁹ However, these examples can do little to help the proponent of (UT). There seems to be far more a priori knowledge than these two cases, and it is not at all clear how reflection on these two cases will motivate the claim that a priori knowledge is in general unrevisable knowledge.

Recent literature contains two further ways of responding to the argument from general revisability. *First*: Casullo's argument turns on the suggestion that we can imagine cases in which we would revise our a priori knowledge. This appeal to imaginability has been challenged. *Second*: Casullo allows, very plausibly, that a priori knowledge is immune to direct defeat. Some have thought that this alone is sufficient to establish negative conclusions about a priori knowledge.

We will deal with these in turn.

Indirect revision and imaginability

Casullo assumes that beliefs in propositions susceptible to logical or mathematical proofs are immune to direct defeat: hence all interest devolves onto whether such beliefs are also immune to indirect defeat. The strategy for showing that all knowledge is revisable turns on the following claims: belief forming processes will be realised by neurophysiological processes; so, for any token process α it is easy to imagine cases in which our warrant for believing in the reliability of α is indirectly defeated by possibly misleading but ostensibly authoritative empirical evidence, which challenges our confidence in the reliability of token neurophysiological processes which realise it.

There are two ways in which we could object to this appeal to the imaginability of defeating cases to motivate the view that knowledge is indirectly revisable: (a) it could be denied that such defeating cases actually are imaginable; or (b) it could be asserted that although they are imaginable, such cases are not (in the requisite sense) possible.

³⁹ See Putnam 1987: 101, also Putnam 1983, chapter six.

Unrevisability

Strategy (a) must do more than simply gainsay the suggestion that a priori knowledge can be indirectly defeated. One possibility might be to hold that supposed cases of defeat are misdescribed. Consider a case in which a token process α produces in S a belief in a logical truth. A defeating circumstance might involve S being told by authoritative neurophysiologists that her ability to reason logically had been impaired, that α was realised by a wholly unreliable brain process: as a consequence, S's warrant for her belief in the logical truth is defeated. Against this it might be maintained that, since logical truths are at stake, in such circumstances S would actually be warranted in retaining her belief in logical truths and assuming that the scientists had somehow got it wrong. But this is question-begging and moreover independently unconvincing. Whatever canons of rationality we employ, there are possible circumstances in which experience is so recalcitrant that we are led to give them up. In such circumstances the very concept of rationality is undermined, *a fortiori* it is not rational for us to maintain any of our beliefs.

Strategy (b) is more interesting. Perhaps there is a way to maintain that defeating cases are imaginable without allowing that they are genuinely possible. Something along these lines is developed by Field, who holds, first, that a notion of a priori knowledge not incorporating an unrevisability requirement is too weak to be of interest but, second, that we must respect the intuition that cases of indirect defeat of a priori knowledge are imaginable.⁴⁰ Field tries to reconcile these requirements by arguing that imaginable defeating circumstances, though *imaginable*, are not genuinely *possible*. He draws an analogy with set theory.⁴¹ It is consistent with our belief that standard set theory is consistent that we can imagine circumstances in which we became convinced that a contradiction could be drawn from it. The fact that we can imagine such a case does not render set theory inconsistent, since it is inconsistent only if a contradiction is *genuinely* (in this case mathematically) possible.

⁴⁰ Field 1998.

⁴¹ *Ibid.*: 5.

Unrevisability

Now, the suggestion that imaginability does not entail genuine possibility is plausible⁴²; the question is whether it finds application in the present case. Misleading evidence is both imaginable *and* genuinely possible: this is indisputable. And misleading evidence is all that is needed to set up cases of indirect defeat of a priori knowledge. Field maintains that we must first find a principled way of distinguishing genuine from misleading evidence, and so maintain that a belief is only "genuinely" revisable if there is no "genuine" evidential situation in which it would be rational to give it up. Field does not suggest how this is to be done, but holds that unless we allow that there is some way of doing it, we trivialise a priori knowledge by allowing too quick an argument for revisability. But his position is unconvincing, for it is not at all clear that it is possible to draw the requisite distinction. For purposes of rational assessment, all evidence is on a par. Thought-experiments involving misleading evidence are set up in just such a way that it seems we have, rationally, to take the misleading evidence into account. Without a principled way of distinguishing genuine from misleading evidence, strategy (b) fails. So Field fails to undermine Casullo's argument that a priori knowledge is indirectly revisable.

Direct revision and scepticism about a priori knowledge

Setting aside Field's argument, we can consider a second way of responding to Casullo. This involves trying to show that the concession that a priori knowledge is not *directly* revisable is still enough to induce scepticism about a priori knowledge. In essentials, this is Devitt's strategy.⁴³ Devitt maintains that Quine's thesis of holism is meant to establish the claim that all knowledge is *directly* revisable. He writes:

Quine's revisability thesis is surely concerned only with [direct revisability]: "no statement is immune to revision" (Quine, 1953, p.43) in that experiential evidence

⁴² Genuine possibility is constrained by the modal and nomological structure of the world, and the claim that the set of imaginable things is a subset of the genuinely possible things is certainly nontrivial. However, I do not need to resolve this issue to make the present point.

⁴³ Devitt 1998.

Unrevisability

might directly bear against *it*. The thesis is simply concerned with the relation between evidence and statement not with the relation between evidence and the view that a particular person (or even a particular community) has thought well about the statement.⁴⁴

Granting for the moment that all knowledge is directly revisable, we can take Devitt to be arguing for a Quinean scepticism about a priori knowledge in the following manner:

- (i) All knowledge is directly revisable
 - (ii) A priori knowledge is not directly revisable
- So, (iii) There is no a priori knowledge.

This argument has power since it turns on a premise - (ii) - which a proponent of revisability such as Casullo is willing to grant: as we saw in the recent quotation, Casullo finds it plausible that having arrived at a belief in *p* via a proof, for example, renders that belief immune from direct revision.

The way to resist this argument, of course, is to reject the revamped first premise. There are two ways of rejecting this premise: one moderate, one radical. The moderate way is to argue that Quine does not manage to sustain the claim that all knowledge is directly revisable. There are two ways in which we might try this. One way would be to point out that Quine's revisability thesis is meant to be supported by his holism, and then argue that Quine does not manage to establish holism. Alternatively, we could grant Quinean holism, but argue against Devitt's interpretation: it is indeed not obvious that Quine is to be read in the way Devitt suggests, and indeed Quine's remark about "pleading hallucination" to reject experiences which tell against a favoured theory suggests strongly that he has failures of cognitive processes in mind, as well as direct defeat.

These moderate options are, I think, promising, and could well be explored by proponents of revisability who wish to reject Devitt's argument. However, I want to consider a more radical way of responding to Devitt's argument: this

⁴⁴ *Ibid.*: 48. What Devitt calls "evidence that someone has not thought well about a statement" is equivalent to what we are calling "indirect defeat".

Unrevisability

response consists of arguing for a very strong denial of premise (i): it is proposed to argue that *no item of knowledge is directly revisable*.

This claim seems problematic: in characterising a priori knowledge as revisable but only indirectly revisable, there is a tendency to assume that this in itself picks out a distinction between a priori knowledge and knowledge in general. This tendency is certainly implicit in the most recent quotation from Casullo: Casullo writes that possession of a proof of a proposition renders belief in that proposition immune to direct defeat, implying that a warrant for a belief derived from something less than a proof leaves that belief still open to direct revision.

But it is wrong to draw a distinction between knowledge and a priori knowledge in point of direct revision: a paradox due to Kripke, which turns on the principle of the closure of knowledge under known entailment, makes this point.⁴⁵ The principle of closure is as follows (writing "K___" for "___ is known"):

$$Kp, K(p \vdash q) \vdash Kq.^{46}$$

Now, clearly, if p is true then all evidence against p is misleading; but if S knows p then p is true. Hence, for any p which S knows, S should be able to reason as follows:

- (1) I know p
- (2) I know (p is true entails all evidence against p is misleading)
- (3) I know that all evidence against p is misleading.

But if S knows that all evidence against p is misleading then there is no evidential situation in which S 's evidence can give S reason to believe $\neg p$. But then, by the definition of direct revisability, S 's belief in p is not directly revisable.

The argument is sound; yet p was arbitrary. Hence it seems that no item of knowledge is directly revisable. Further, the only principle used to establish this

⁴⁵ This setting up of the paradox is drawn from Lewis 1996.

⁴⁶ The principle is discussed further in the following chapters.

Unrevisability

conclusion was the principle of closure of knowledge under known implication. We will return to this principle later; suffice for now to say that adoption of the principle seems motivated by very general considerations about the concept of knowledge. The claim that no item of knowledge is directly revisable, then, seems to be central to the concept of knowledge, and I think that despite its paradoxical appearance, we must accept it. The appearance of paradox can be lessened, and perhaps dispelled, once we have gone into more details about the general theory of knowledge.

Conclusion

We have found no difference between knowledge in general and a priori knowledge in point of either direct or indirect revisability. The naturalistic theory of knowledge, which links the warrant for a belief with the causal ancestry of the belief, renders all knowledge open to social challenges by experts⁴⁷, and hence renders all knowledge indirectly revisable. To this extent, Kitcher is right to take cases of social challenge to knowledge as showing that mathematical knowledge is revisable. Where he fails is in giving us any reason to think that a priori knowledge is unrevisable knowledge. The revisability of a priori knowledge therefore seems to be something that naturalistic theories of a priori knowledge will have to accept. However, the claim that knowledge is immune to *direct* defeat seems to have been established on the basis of very general considerations of the concept of knowledge: it is therefore something that must be accepted by all theorists of knowledge, naturalistic or otherwise. Therefore we cannot even adopt the very weak position, that while a priori knowledge is only indirectly revisable, knowledge in general is either directly or indirectly revisable. All knowledge is indirectly revisable; and no knowledge is directly revisable.

It seems, then, that we should not appeal to the concept of unrevisability in explaining the "experience-independence" of a priori knowledge. This leaves us

⁴⁷ All - bar the two anomalous cases noted on page 117.

Unrevisability

with an account of experience-independence in terms of the minimal notion alone, or that notion augmented by the stipulation that a priori warrant is infallible. The following chapters argue for a position as regards infallibility analogous to the one taken here on revisability: it will be claimed that no difference can be drawn between a priori warrant and warrant in general in point of fallibility. Having established this, and dealt with a problem which arises, we turn in the final chapter of the thesis to discuss an account of experience-independence in terms of the minimal notion alone.

CHAPTER SIX

Infallibility

The third option available to us in explaining "experience-independence" was to appeal to the concept of infallibility. On this account, experience-independence is analysed in terms of the notion of nonexperientiality combined with the requirement that the experience-independent warrant never attach to a false belief. We will say that a property of beliefs is *infallible* just in case necessarily every belief that has it is true. That is, we adopt the following definition of infallibility (for property of beliefs ϕ):

$$\phi \text{ is infallible iff } \forall p(\phi p \rightarrow p \text{ is true}).^1$$

To say that a priori warrant is infallible is, then, to say that no false belief can be warranted a priori.

On Kitcher's definition, the infallibility of a priori warrant is required directly by clause (3c).² However, the argument that Kitcher offers for adopting clause (3c), and hence the requirement of infallibility, is less explicit than his argument for adopting unrevisability. In favour of clause (3c), he cites an intuition:

The intuition is that a priori warrants must be ultra-reliable: if a person is entitled to ignore empirical information about the type of world she inhabits then that must be because she has at her disposal a method of arriving at belief which guarantees *true* belief. (This intuition can be defended by pointing out that if a method which could produce false belief were allowed to override

¹ We might wonder whether the modal requirement could be dropped, and whether only the material truth of the conditional $\phi p \rightarrow p$ is required for infallibility. To see that the stronger version is needed, consider a blackboard B with a true proposition written on it; then, the predicate "___ is written on blackboard B" is infallible. Intuitively, this is not what we want. Roughly speaking, the difference between the non-modalised and the modalised version of the infallibility requirement is that between "X didn't produce a false belief" and "X couldn't produce a false belief".

The *grade* of necessity appealed to has not been specified. This problem will be deferred until the discussion of *relevance* in the following chapters.

² See chapter four.

Infallibility

experience, then we might be blocked from obtaining knowledge which we might otherwise have gained.) In my analysis, the intuition appears as (3c).³

There are two proposals here; the first, coming before the parenthesis, is of little help as it stands: it amounts merely to an intuition in favour of infallibilism. But this quotation from Kitcher comes in the context of a critique of an argument that only necessary truths are knowable a priori. Roughly stated, the argument Kitcher criticises is as follows:

Assume S knows p a priori. Her knowledge is independent of her experience. Hence S can know p without any information about the kind of world she inhabits. So, necessarily p .⁴

Kitcher writes that this argument turns on a misconstrual of the intuition, cited above, that a priori warrant is infallible. This, perhaps, suggests that Kitcher has in mind the following argument for the infallibility of a priori warrant, which is structurally similar to the rejected argument just sketched. For a priori warranting process α :

- Assume (i) α could warrant a false belief
- so, (ii) in some possible world, α does warrant a false belief
- so, (iii) in order to find whether α warrants a false belief in this world, we have to inspect our world
- so, (iv) in order to use α , we have to inspect our world
- so, (v) α is not an a priori belief forming process
- so, (vi) no a priori belief forming process could warrant a false belief.

This argument, however, is unconvincing. The weak point is step (iv). It is weak because there is no reason to think that the epistemic efficacy of a belief forming process depends on our having checked to see if it is appropriate. Quite the

³ Kitcher 1983: 29-30.

⁴ *Ibid.*: 29. His critique of this argument, which does not concern us here, turns on counterexamples to it such as the proposition expressed by an utterance of "I exist". Such propositions are arguably knowable a priori; they are also contingent.

Infallibility

contrary: the guiding idea behind Kitcher's psychologistic account of knowledge seems to be a shift to a view that the epistemic efficacy of our belief forming processes is independent of whether we know them to have the appropriate properties. After all, we do not, now, know what the appropriate properties are, yet we do possess knowledge. So step (iv) is unacceptable.

A second argument for the infallibility of a priori warrant is proposed in the parenthesis towards the end of the quotation. Here Kitcher defends his claim that a priori warrant is infallible by suggesting that if false beliefs could be warranted a priori then we might be "blocked from obtaining knowledge which we might otherwise have gained". But this suggestion has weight only if we are willing to assume that a priori knowledge is unrevisable knowledge. In the case of ordinary knowledge and ordinary warrant, we tend to think that we can allow that some false beliefs get warranted, for we are also inclined to think that other processes will be available to identify and correct warranted false beliefs when they occur. The operative assumption here is that ordinarily warranted beliefs are revisable, and it is only because Kitcher thinks that a priori knowledge is unrevisable knowledge that he thinks that a priori warrant must be infallible. Once we drop the unrevisability assumption, we have no reason to think that a fallible notion of a priori warrant would commit us to the possibility of false beliefs that we were unable to correct, and so no reason to adopt infallibilism.⁵

No further argument seems to be forthcoming for the infallibility of a priori warrant. Indeed, many authors take the *fallibility* of a priori warrant to be easily established.⁶ Thus, in Casullo's case B, in the previous chapter, it is assumed that perceptual warrant is fallible, and Casullo takes there to be a significant analogy between case B, and case A, where Mary arrives at a false belief through a process of reflective thought, such that it is obvious that a priori warrant is also fallible.

In the same vein, Summerfield considers an expert mathematician, a highly reliable prover of mathematical truths.⁷ Summerfield takes it that on the rare

⁵ The previous chapter discussed a related point.

⁶ Casullo 1988, Summerfield 1991.

⁷ Summerfield 1991: 55-6.

Infallibility

occasion where the mathematician executes a flawed proof, and so comes to believe in a falsehood, the process of following that flawed proof must be taken as an instance of a generally reliable process type. Hence, she concludes, since it was produced by a reliable process, the false mathematical belief is warranted.

So Casullo and Summerfield take it that a priori warrant, like warrant in general, is fallible. But they are wrong to do so. This chapter will argue that all warrant is infallible; it will follow that a priori warrant is infallible. This conclusion will be established not on the basis of any specific considerations of the concept of a priori knowledge, but through reflection on the general theory of knowledge. By way of introducing the argument that warrant is infallible, it will be useful to begin by emphasising the difference between warrant - which we are using as a technical term - and the more intuitive notion of justification.

Justification

Following Kitcher, we are using the term "warrant" as a placeholder for whatever property X solves the equation: $\text{Knowledge} = \text{True Belief} + X$. Some authors use the term "justification" in this capacity, but this is unhelpful, since we already possess strong intuitions about the concept of justification. These intuitions, however, are not wholly determinate (as we will shortly see) and have been deformed by the pressure brought to bear on the concept of justification as a result of its employment in what is called the "traditional" analysis of knowledge. On the traditional analysis, knowledge is taken to be a combination of truth, belief and justification. That is:

(JTB) S knows p if and only if p is true, S believes p and S 's belief in p is justified.

p must be true since it is intuitively inconsistent to assert simultaneously that S knows p and that p is false: this is to say that knowledge is factive, only truths can be known. The assumption that p must be also believed is central to the project of

Infallibility

giving a reductive analysis of knowledge. Typically, the state of knowing p is taken to be a hybrid of a mental and a metaphysical state.⁸ The metaphysical aspect of the state is given by the requirement that p be true; the fundamental mental aspect is that p be believed. Believing is assumed to be more fundamental than knowing, and the task of understanding knowledge is taken to be the task of showing what extra conditions must be met for a belief to be known.

The conditions of truth and belief are not, however, jointly sufficient for knowledge since, uncontroversially, there can be true beliefs that are not known. Plato's *Theaetetus* contains an early demonstration of this: a skilled advocate may convince a jury of the rightness of his case through force of oratory alone; but while the jury may through this method come to believe truly in the rightness of the advocate's case, they do not possess knowledge that he is right.⁹ True belief needs to be qualified by something else to give a full analysis of knowledge, and the traditional model suggests that this third factor is *justification*. This requirement is acutely difficult to clarify. There are probably (at least) three notions of justification which can lay some claim to being intuitive. The fundamental distinction is between *regulative* and *nonregulative* conceptions of justification.¹⁰ A regulative norm provides rules and precepts which an agent can use to guide her actions; a nonregulative norm provides standards against which actions can be judged, but does not supply guidelines for the agent to follow from the first-person point of view. A regulative standard of epistemic justification governs what a subject thinks she ought to believe given her evidence; a nonregulative standard of epistemic justification is used to judge whether the agent is right in believing something. It is not part of the concept of nonregulative justification that the agent be aware of which of her beliefs are justified in the nonregulative sense, or what the standards are for assessing them. Roughly, nonregulative justification concerns a privileged observer's assessment of whether S is right in believing something.

A second distinction, which overlaps the first, is between *internalist* and *externalist* notions of justification. A conception of justification is internalist just in

⁸ See for example Ayer 1956: 14ff., but cf. Williamson 1995.

⁹ *Theaetetus*: 201a-201c. See Burnyeat 1990

¹⁰ Goldman 1986: 25.

Infallibility

case, (for all p) if S 's belief in p is justified then S knows it is justified; a conception of justification is externalist otherwise.¹¹ The second distinction should not be conflated with the first. A norm could be regulative without being internalist: one may be able to follow a rule without being able to formulate it precisely, and so without knowing in every case whether one is following it.¹²

I am inclined to think that our intuitions about justification are regulative and internalist¹³, and that the impulse underlying the traditional analysis of knowledge is captured most nearly by the following definition of justification:

S 's belief in p is *justified* iff S has good reasons for believing that p is true.

A "good reason" is *another justified belief* which S , if competent and rational, would on reflection judge makes p likely to be true. The requirement that a good reason be one that makes S *think* that p is likely to be true, as opposed, say, to one which makes p objectively likely to be true, stems from the regulative role of justification: justification in the intuitive sense is primarily to do with whether S has done the right thing *by her own lights* in believing p - whether S is epistemically responsible in believing p , or whether she cannot be blamed for believing it. It follows that S can have good reason to believe something false, and hence false beliefs can be justified: justification is *not* an infallible property of beliefs.

The suggestion that a good reason is a *belief* can be defended by observing that reasons are essentially things which influence a subject's actions, and the mere fact that an event occurred cannot influence S 's actions unless S also believes that it occurred.¹⁴ The suggestion that a good reason must be a *justified* belief can be defended by considering how we would react in a case in which S , on being asked for her justification for her belief in p , cites another belief q , but that on being asked

¹¹ This is not the only way of formulating the distinction, but it captures the general idea. See BonJour 1985 for an account given in terms of "cognitive accessibility".

¹² Cf. Wittgenstein 1953: §§208ff; Williamson 1996a, 1997.

¹³ I will use the term "justification" to refer to the regulative internalist sense, unless otherwise indicated.

¹⁴ There is an attenuated sense of "reason" in which this is not so. In this attenuated sense the fact that Catfish Keith has arranged a gig in Edinburgh can be a reason for me to be cheerful even though I haven't yet heard about the arrangement, and so neither know nor believe that he is going to perform. But this sense of "reason" seems to be nonregulative, and I will not consider it further.

Infallibility

for her justification for her belief in q can say nothing. I suggest that our intuitions in this case are that S 's belief in p was not justified.¹⁵

This last requirement, that the justifying beliefs be themselves justified, launches us on a regress. If any beliefs are to be justified at all, either the chain of regress must loop around on itself, with all beliefs entering indirectly into their own justification, or the regression must terminate with a class of *basic beliefs* which can confer justification without themselves being justified. The introduction of the concept of a basic belief invites a specification of the sorts of objects which these beliefs are about. Many authors have thought basic beliefs to concern a special class of objects, variously called "sense data", "sense contents", "simple ideas" etc., which are not misidentifiable. Others have allowed that basic beliefs concern less abstruse entities: for example, middle sized physical objects. This is not, however, a problem that we need to resolve here.

Adoption of the first of these positions as to the ultimate source of justification gives rise to the position known as *coherentism*. A system of beliefs is justified if it forms a coherent system.¹⁶ Coherentism is not currently widely endorsed, for the coherentist seems to leave it open that the justification of our beliefs could float free, and so be as arbitrary as the beliefs of a paranoid or conspiracy theorist.¹⁷ So long as the paranoid's system of beliefs is coherent, it seems that the coherentist must say that they have overall justification; but this is an uncomfortable position.

To adopt the idea that the ultimate source of justification derives from basic beliefs is to adopt *foundationalism*. However, on the present conception of justification, the adoption of foundationalism about justification raises serious problems for the JTB model of knowledge. For it seems plausible that basic beliefs, at least in certain favourable cases, can be items of knowledge.¹⁸ But basic beliefs

¹⁵ This intuition is what drives Hume's problem of induction. See for example discussion in Bird 1998.

¹⁶ See for example Bonjour 1985.

¹⁷ This objection parallels the standard "Plurality Objection" levelled at coherence theories of knowledge. See for example Dancy 1985: 113-4.

¹⁸ Illustrating this point requires us to take a temporary stand on the nature of the objects of basic beliefs. Assuming that basic beliefs concern middle sized physical objects: where S has a true basic belief that middle sized physical object o is before her, and where the prevailing epistemic conditions are favourable, it seems intuitively right to say that S knows that o is before her. (It also seems that on

Infallibility

are, on this conception of justification, not justified. Hence we seem to have a class of cases of knowledge without justification: and the proposed equivalence in (JTB) fails from left to right: justified true belief is not necessary for knowledge.

Faced with this, some authors have suggested that it is illegitimate to ask what is the justification for basic beliefs. Clarke, for example, suggests that to ask someone for their justification for their belief that their memory is reliable is to ask a question that makes no sense.¹⁹ But this is peculiar: there is nothing grammatically wrong with the sentence: "what justification do you have for believing that your memory tends to produce true beliefs?"

A better way of responding to this challenge to the JTB model would be to modify our notion of justification. One way of doing so would be to allow that *experiences*, as well as reasons, can justify beliefs. Adopting this view would enable us to claim that basic beliefs are justified by the experiences which engender them, and this is a fairly common foundationalist stance on the concept of justification, although it is not clear whether we should treat this notion of justification as a technical or as an intuitive concept.

However, the JTB model of knowledge is not thereby saved, for as Gettier demonstrated, it fails also to give a sufficient condition for knowledge.²⁰ Gettier inaugurated the project of providing counterexamples to analyses of knowledge; the project is associated with him to such a degree that the term "Gettier case" has become synonymous with "counterexample to a mooted analysis of knowledge". A typical Gettier case is:

- D S is in a room with two other people. She has known one of them for years, and knows that he has always loved Ford cars, that he has always bought them over other cars, that he has never expressed any interest in driving a non-Ford car, etc. S also knows that yesterday this friend arrived at work in his own Ford car, and that he drove into work today (though she didn't see him arrive). S has, therefore, a justified belief that her friend owns a Ford car: but in fact he does not, having sold it just that morning. S does not own a Ford car. However, the second person, who is a stranger to S, does own a

Williamson's conception of evidence, basic beliefs are known since they compose part of their possessor's evidence (Williamson 1997).)

¹⁹ Clarke 1963.

²⁰ Gettier 1963.

Infallibility

Ford car. Thus, S has a justified and true belief that *someone in the room owns a Ford car*.

S does not know that someone in the room owns a Ford car. Hence, having a justified true belief is not a sufficient condition for having knowledge. Moreover, the relaxed notion of justification does not identify a necessary condition for knowledge either: for even allowing experiences to justify beliefs, we do not seem to be able to give a noncircular explanation of how we are justified in relying on perception or memory.²¹ Since we can get knowledge through these channels, though cannot justify our reliance on them, we will have cases of knowledge without justification. The claim that modified-justification is unnecessary for knowledge can be backed up with specific examples: a terrified student sits down to an exam, his mind a blank, but is able to produce the right answers; in his state of confusion he is not able to give reasons for his right answers, but nevertheless it seems right to say that he knows them.

Another proposed modification to the concept of justification is to link the justification of a belief to the *metaphysical* probability of that belief being true. That is, to require that good reasons to entertain a belief are reasons that make the belief objectively likely to be true, as opposed to ones that make S think it is likely to be true. This is clearly an externalist view of justification: since S will not always be in a position to determine whether her reasons do confer metaphysical probability on her belief over and above subjective likeliness to be true, she will not always be in a position to know which of her beliefs are justified. However, it seems that the proposed modification to the concept of justification will not help to shore up the JTB model of knowledge in the face of case D. The proponent of the externalist version of the JTB model would propose that on the externalist theory of justification, it would turn out that S's true belief that someone in the room owns a Ford car is not in fact justified, and that therefore case D does not present a counterexample to the JTB model. But this is implausible: S's belief that someone in the room owns a Ford is based on her having a great deal of positive evidence, and

²¹ Lewis 1996: 551.

Infallibility

no negative evidence, for this belief. To deny that S is (externalistically) justified in her belief is to deny that inferring p from a preponderance of unequivocal evidence makes p likely to be true, and this is surely controversial. However, the externalist notion of justification may constitute a necessary condition for knowledge, and we will discuss Goldman's development of this idea in the following chapter.

These considerations of the concept of justification have been intended to illustrate the difficulties faced in giving an analysis of knowledge. The project of the theory of knowledge since Gettier has been to find out what should be substituted for "justification" to give an analysis of the concept of knowledge. To prescind from confusions about what the concept of justification really involves, it is best to introduce a new technical term and to conduct the investigation into the nature of knowledge in terms of it. This is, of course, the concept of warrant.

Warrant

As already noted, "warrant", as we are using it, is whatever property solves the equation: Knowledge = True Belief + X. The use of the term "warrant" - a cognate of the term "justification" - is meant to respect the intuition behind the traditional appeal to justification, that a true belief is known if it is believed for the right reasons. The preceding considerations indicate that strong subjective reason is not good reason enough: it is not the case that a belief is warranted just in case it is justified. The naturalistic shift endorsed by Kitcher is to interpret "right reasons" not epistemically but metaphysically - as with the externalist conception of justification, warrant will be connected with *metaphysical* as opposed to epistemic probability. Roughly speaking, a proposition p is metaphysically probable just in case it is objectively likely to be true relative to the way in which it came to be believed.

The crucial property of warrant is brought out by considering the inferential structure of the Gettier case D. The structure of case D is as follows: S has a justified false belief of the form Fa ; Fb is a true proposition which S has no way of knowing is

Infallibility

true; and we allow that justification for belief in a proposition is transmitted across to the logical consequences of that proposition. As, in classical quantificational logic, any proposition can validly be weakened according to the inference schema $Fa \vdash \exists xFx$, S's justification for her false belief in Fa transfers over to her belief in $\exists xFx$; the proposition $\exists xFx$ is true, because Fb is true, but S's justified true belief that $\exists xFx$ is not an item of knowledge. From this we can identify two relevant assumptions underlying the Gettier case:

- (i) justification can attach to false beliefs
- (ii) justification is transmissible across logical entailment.

(ii) must be handled with some care, but becomes very plausible when suitably restricted. The principle of transmission of justification can be formulated as follows (writing "J___" for "S is justified in believing ___"):

$$J\alpha, (\alpha \vdash \beta) \vdash J\beta.$$

So formulated, the principle is implausible; for, let α and β be mathematical truths such that α entails β , where α is very simple and β is immensely complex.²² It may be that while S is justified in believing in α , S is not only unaware of, but in fact incapable of grasping, the proposition $\alpha \vdash \beta$. It is then highly implausible that S is justified in believing in β . However, restricted to cases where S can grasp α and β and knows the relation between them - as is the case in the Gettier case under discussion - the principle gains plausibility. This plausibility derives from the affinity between the restricted formulation and the principle of the *closure* of knowledge under known implication. The principle of the closure of knowledge is, as we have seen:

$$K\alpha, K(\alpha \vdash \beta) \vdash K\beta.$$

The restricted principle of transmission is:

$$J\alpha, K(\alpha \vdash \beta) \vdash J\beta.$$

²² See Dretske 1970, Stine 1976.

Infallibility

The principle of closure is very plausible, though it has been contested, chiefly by Nozick and Dretske; their doubts about closure, however, are not now widely shared.²³ Suffice to say that (ii) needs reformulating: the second principle underlying Gettier cases is as follows:

(ii') justification is transmissible across known logical entailment.

It is because the intuitive concept of justification makes (i) and (ii') true, that we can generate counterexamples to the suggestion that knowledge can be analysed in terms of truth, belief and justification. Generalising: any account of the concept of warrant which makes true the appropriate reformulations of (i) and (ii') will be susceptible to Gettier cases structurally isomorphic to D. Hence, whatever account of warrant we endorse, it cannot be that that account makes true both the following propositions:

(wi) warrant can attach to false beliefs

(wii) warrant is transmissible across known logical entailment.

The plausibility of the principle of closure makes it extremely likely that whatever account of warrant we develop, it must be transmissible across known logical entailment - that is, it must make (wii) true.²⁴ It follows, then, that no concept of warrant respecting (i) will escape Gettier cases isomorphic to case D. The idea is this: since any conception of warrant which makes (wi) and (wii) true will be susceptible to Gettier cases, and (wii) is very plausible, we cannot - if we want to avoid Gettier cases - give an account of warrant which makes (wi) true. So we must deny that warrant can attach to false beliefs.

²³ Dretske 1970, Nozick 1981. Closure is defended in, for example, Stine 1976, Cohen 1988, DeRose 1995, Lewis 1996. See chapter eight.

²⁴ To this it could be objected that the concept of warrant, as we are using it, is a technical term, and so our intuitions about whether it is closed under known logical entailment do not apply. It would be difficult to give a fully worked out response to this objection; however, it seems likely that the property of warrant must turn out to be some form of natural property and it is not easy to see how it could fail to be closed under known implication. For the purposes of this thesis, I will assume that it is closed.

Infallibility

As before, a property of beliefs is *infallible* just in case necessarily only true beliefs can have it. The conclusion to be drawn from the present discussion is this: *warrant is infallible*. This conclusion is also urged by Zagzebski, whose account generalises on the considerations given above.²⁵ Zagzebski's account of the general structure of Gettier cases does not require any inferences to be made, and so does not depend on any account of principles of transmission of warrant. Given an account of warrant such that false beliefs can be warranted, we construct a counterexample to a definition of knowledge employing that notion of warrant as follows: (i) allow subject S to have a false belief in *p*; (ii) allow that S's false belief is warranted with sufficient strength such that a true belief so warranted would be known; (iii) ensure that the falsity of *p* does not depend on any systematically describable feature of the situation which might be ruled out by features of the analysis of knowledge independent of warrant (for example: "bad epistemic luck"²⁶); (iv) emend the case by allowing that, by an independent case of bad luck, *p* is actually true. Hence, we have a case of warranted true belief which is not knowledge.²⁷

Warrant is infallible. It follows that the initial formulation of the definition of knowledge in terms of warrant has a redundant clause. Since possession of warrant for a belief guarantees the truth of that belief, we need not devote a separate clause of the definition of knowledge to the requirement of truth. The concept of warrant needed to explicate the concept of knowledge is such that:

S knows *p* if and only if S has a warranted belief in *p*.

²⁵ Zagzebski 1994.

²⁶ Paradoxically, having identified bad epistemic luck as the feature that prevents us from developing an analysis of knowledge, Zagzebski tentatively proposes an account of knowledge essentially as follows: S knows *p* iff S believes in *p*, *p* is true, S's belief is warranted, and S is not epistemically unlucky. We should take Zagzebski's proposal more as an admission of failure than as a proposed analysis of knowledge.

²⁷ Zagzebski 1994: 69.

Infallibility

This result is extremely significant, but also problematic. It is significant because if all warrant is infallible, then a priori warrant, as a subset of warrant in general, is also infallible. So Kitcher is right that a priori warrant is infallible: but the reason why he is right totally undercuts the project of explaining "experience-independence" in terms of the concept of infallibility. All warrant, whether a priori or not, is infallible. Though Casullo and Summerfield are wrong to suppose that the fallibility of a priori warrant is an easy consequence of the fallibility of warrant in general, they are right to think that a priori warrant is no more or less reliable than warrant in general. No difference can be drawn between a priori warrant and warrant in general in terms of whether false beliefs can be warranted. So, an account of "experience-independence" which incorporates infallibility will not help us characterise a priori knowledge.

The result is problematic because it is not clear how a naturalist theory of knowledge will respect it. It is natural for the naturalist to appeal to the method of production of a belief to explain its epistemic status: but all our sensory and inferential processes produce occasional false beliefs, so none of them is infallible. This is not yet a problem, since we must distinguish the idea of warrant being infallible from a belief forming process being infallible: given that our normal processes are fallible, there must be more to being a warranted belief than just having been produced by one of the normal sensory or inferential processes. Some extra condition or conditions must also be filled, and we should note that whatever these extra conditions are they will have to be shown to be scientifically acceptable if the concept of warrant - and, by extension, the concept of knowledge - is to be naturalistically admissible. We will consider whether this condition has been met after looking at the naturalist's proposed solution to the problem of the nature of warrant: that is, the adoption of reliabilism.

CHAPTER SEVEN

Goldman's reliabilism

Kitcher's strategy for explaining a priori knowledge has two parts. One part involves giving an account of experience-independence; the other involves giving a naturalistic theory of knowledge in general. Solutions to these are intended to be combined to give an account of experience-independent knowledge - that is, a priori knowledge. We have seen that these two sub-projects cannot be carried out independently, for results from the general theory of knowledge affect what account we can give of experience-independence. But in exploring how these projects interact, we have uncovered a problem for any theory of knowledge in general: this is that warrant, whatever that property may be, cannot attach to false beliefs. Chapters seven and eight assess the prospects for giving a naturalistic account of infallible warrant.

Kitcher, as well as many authors who have followed him in giving naturalistic accounts of a priori knowledge, has turned to *reliabilism*, and in particular Goldman's writings on reliabilism, to find an account of warrant and knowledge in general.¹ The general idea behind reliabilism can be illustrated by considering the following case:

- E S uses astrology to form some belief *p*. *p* is in fact true, and S has complete subjective justification for her belief, having been convinced of the efficacy of astrology through being raised in an community where astrology was common, having seen apparently reliable results many times before, being versed in the underlying pseudo-theory, etc.

S does not know *p*. The obvious explanation for S's failure of knowledge in case E is that there is something wrong with the belief forming process S used to form her belief. The astrological divination process does not confer any real metaphysical probability of truth on S's belief. A belief can be said to be metaphysically probable just in case it is objectively likely to be true relative to the

¹ Cf: Kitcher 1983, 1992; Casullo 1988; Summerfield 1991.

Goldman's reliabilism

way that it was brought about. This is equivalent to saying that the astrological divination process is *unreliable as a method of producing true beliefs*. A belief forming process is reliable just in case it produces a high proportion of true beliefs.² Generalising: if S's belief in *p* is produced by an unreliable belief forming process, S does not know *p*. Contraposing: if S knows *p* then S's belief in *p* is produced by a reliable belief forming process. This gives us a necessary condition for a belief forming process to confer warrant on the beliefs it produces: a belief forming process α confers warrant on the beliefs it produces only if α is reliable.

As we have already mentioned, an account of this sort is generally thought to be what Kitcher has in mind when he appeals to "appropriate" production. The *reliabilist* is one who seeks an analysis of warrant in terms of reliability. Reliabilist theories fall into two groups: on a *process-reliabilist* theory, a belief *p* produced by a belief forming process α is warranted just in case α is reliable; on a *tracking* theory, the reference to processes is left out, and a belief *p* is considered warranted for S just in case there is a high degree of correlation across a range of possible worlds between S's believing in *p* and *p*'s being true.³ Unadorned tracking theories are not generally considered to be promising, and we shall follow Kitcher and others in concentrating on reliable process theories.⁴ Process-reliabilist theories of knowledge are the most plausible implementation of naturalised epistemology, and the best worked out process-reliabilist theory is generally taken to be Goldman's. Goldman's general intention throughout his epistemological work is to reduce epistemic concepts such as knowledge and justification to nonepistemic concepts such as reliability, belief and truth.⁵

The present challenge is for the reliabilist to give an account of warrant on which warrant is infallible. The basic problem they have to address is that it is

² Authors do not tend to take a stand on just what proportion of truths is required for reliability, but, intuitively, the appeal to reliability imports a proportion significantly greater than 50%, but less than 100%.

³ The taxonomy is from Goldman 1986.

⁴ The general problem is that tracking in counterfactual situations can be gerrymandered; see Goldman's "smiling stranger" example (Goldman 1986, ch.2); see also Fogelin's discussion of Dretske's subjunctive account of sufficient reasons (Fogelin 1994, ch.4). Nozick, the author of an influential tracking theory, recognises the problem and adds a reference to processes to his tracking theory (Nozick 1981).

⁵ See esp. Goldman 1992: 105-7.

Goldman's reliabilism

consistent with a belief's being produced by a merely *reliable* belief forming process that the belief be false. If the reliabilist is to give a full analysis of knowledge they will have to introduce into their definition a compensating mechanism such that, on the occasions that a reliable belief forming process produces a false belief, that false belief is not warranted. In the following discussion we will look at three candidates for such a mechanism. The first is a requirement of a causal connection between the subject's belief and the fact that the belief is about; the second is a requirement that a reliable process be one which is able to discriminate between the actual situation and relevant counterfactual situations; and the third is a requirement that the belief to be warranted be justified (in some sense of "justification", possibly not the intuitive sense, as we will see). The second of these has the best chance of success, and discussion of it will continue into chapter eight. However, even this, the best of the three options, faces serious problems: for it is not clear whether the requirement can be cashed out in a way that is naturalistically acceptable.

We begin by looking at the causal requirement. In fact, the causal theory of knowledge is really a prototype reliabilist account and has no serious chance of success. We discuss it in order to introduce the main problems that reliabilism must address.

"A Causal Theory of Knowing"

The causal theory of knowledge⁶, which Goldman advanced prior to the development of his more complex reliabilist accounts, requires that a causal connection obtain between the belief that is an item of knowledge and the fact that it concerns.⁷ Goldman's theory is a development of Grice's suggestion that perceptual knowledge depends on the presence of a causal link between the subject and what the subject observes: it is the presence of this link that confers warrant on the subject's beliefs about her perceptual world.⁸ Goldman generalises Grice's

⁶ Goldman 1967.

⁷ Goldman's theory is couched in terms of fact-causation, and I adopt his framework for ease of exposition. This will not affect the conclusions we will draw.

⁸ Grice 1961.

Goldman's reliabilism

account by attempting to develop the underlying idea into an account of other sorts of knowledge, such as knowledge through memory and inference. Goldman's initial formulation of the idea is as follows; discussing an example structurally identical to case D, he writes:

...one thing that seems to be missing in this example is a causal connection between the fact that makes p true (or simply: the fact that p) and [S's] belief of p . The requirement of such a *causal connection* is what I wish to add to the traditional analysis.⁹

This suggests that Goldman adopts a definition of knowledge as follows:

S knows p if and only if: p is true, S believes p , S's belief in p is justified, and there is a causal connection between the fact p and S's belief in p .

If by "justification" Goldman means the intuitive concept of justification, this definition will not work: where p is a basic belief there will be situations where S knows p but does not have a justified belief in p . Later in the paper, however, Goldman gives a different formulation. Slightly, and harmlessly, recast, it is:

S knows p if and only if the fact p is causally connected in an "appropriate" way with S's believing p ¹⁰

where "appropriate" remains to be explained. This formulation does not require that a subject has a justification for her beliefs and so is not vulnerable to objections based on considerations regarding the justification for basic beliefs. Goldman assumes that knowledge through perception, memory, inference, etc., can all be reconstructed as appropriate causal processes. Note that a *direct* causal link - such that a belief is warranted just in case it is one of the causal effects of the fact that it concerns - cannot be what is meant by "appropriate" since there can be knowledge of the future, and it is unlikely that future events can have causal effects on the present. Where p is a proposition about the future, Goldman takes it that S can

⁹ *Ibid.*: 358.

¹⁰ *Ibid.*: 369.

Goldman's reliabilism

know p where her belief in p is caused by facts in the causal ancestry of p - that is, S 's belief in p , and the fact p , have a common cause. For example, seeing a missile flying towards a building, S forms a true belief that she is about to experience an explosion: this belief counts as knowledge on Goldman's view since it is caused by a fact that is (partially) causally responsible for her experience of the ensuing explosion - namely, the fact that the missile was flying towards a building. "Appropriate" causal connection therefore encompasses indirect connection.

Goldman's 1967 account passes the requirement of infallibility: for it is not possible to be causally connected - appropriately or otherwise - with a fact that does not obtain. Hence, on Goldman's 1967 account of what it is for a belief to be warranted, if a belief is warranted it is true. However, Goldman's 1967 account is in other respects unsatisfactory. For one thing, it seems unable to explain a priori knowledge. Benacerraf's objections apply squarely to an account of a priori knowledge couched within a causal theory of knowledge, for it is very hard to see how the typical objects of a priori knowledge can enter into causal relations.

The causal theory faces further objections. Consider the following case:

- F S has a certain belief p about her early childhood, but her parents tell her that her belief in p is false, and that her memory of the occurrence of p is mistaken. Despite the attempted deception, S continues to believe in p . S 's belief is true and her memory is working perfectly.¹¹

S does not know p , though were it not for her parents' attempted deception she would know it: her knowledge has been undermined by the attempted deception, which leaves her belief in the trustworthiness of her memory unjustified.¹² Since S 's memory is in fact working perfectly, she is appropriately causally connected with the fact about her childhood, and yet does not have knowledge. So there must be more to knowledge than simple causal connection; sometimes, at least, knowledge can be defeated by lack of justification.

¹¹ Based on an example from BonJour 1985.

¹² Note that this example does not show that justification is a necessary condition for knowledge: we have already excluded this possibility by offering counterexamples to the JTB model of knowledge. The example does show that knowledge is in *some* cases sensitive to justification, however: just how sensitive, we do not know.

Goldman's reliabilism

A further problem with Goldman's 1967 account is illustrated by the following case:

- G S is driving along a road, and believes that she sees a barn, based on the fact that she seems to see the front of a barn that she is passing, and that she has no reason to believe that visual conditions are abnormal. Visual conditions are in fact normal, and in fact there is a real barn before her. However, unknown to S she is in a region of the country where many fake barn façades have been erected, such that if S had been looking at a barn façade she would still have believed she was looking at a barn, and it is pure luck that S is now looking at a barn.¹³

S does not know she sees a barn. This seems to be due to a combination of two factors: first, that it is just good luck that the object in front of S is in fact a barn, since the majority of apparent barns in the area are mere façades. Second, that if there had been a fake barn in front of her she would still have believed that she saw a real barn. The fact that her belief is only true by luck undermines her knowledge, even though her visual belief forming processes are functioning normally and so, it seems, are reliable.

The 1967 account does not have the resources to explain why S does not know in case F or case G. Let us say that in case F S's knowledge suffers *internal defeat*; this terminology is suggested by the intuition that the feature of the case which defeats S's knowledge is something internal to S's cognitive perspective.¹⁴ In case G we will say that S's knowledge suffers *external defeat*: the feature of case G which defeats S's knowledge is something outside her cognitive perspective. Goldman's 1967 account, then, fails to give an adequate account of internal or external defeat of knowledge. What we need, then, is an account of warrant that

¹³ This example discussed by Goldman 1976, and attributed originally to Carl Ginet.

¹⁴ The idea of an object being "internal to S's cognitive perspective" is closely related to the possibility of S being directly aware of that object. As Bonjour (in his entry on "Externalism/Internalism", in Dancy and Sosa 1993) points out, it seems to be neither necessary nor sufficient for an object to be internal to S's cognitive perspective that it be a mental object. Not necessary, since on direct realist views of perception nonmental features of the world can be the objects of direct awareness; not sufficient, since on some views of the mental there can be mental objects which are not objects of direct awareness.

Goldman's reliabilism

not only preserves its infallibility, but also gives rise to a definition of knowledge which is immune to counterexamples based on internal and external defeat.

The development of Goldman's epistemological theory from the causal theory to the account in *Epistemology and Cognition* has two strands: one, addressed in the paper "Discrimination and Perceptual Knowledge", involves modifying the causal account to deal with external defeaters; the other, picked up in "What is Justified Belief?", involves modifying the causal account to deal with internal defeaters. The two strands are reconnected in the account in *Epistemology and Cognition*. We will discuss the 1986 account after a survey of the prior papers, which will highlight the problems which the 1986 account must overcome. Our intention is to see whether the modifications Goldman proposes in these papers will provide us with the mechanisms needed to develop a working reliabilist account of infallible warrant.

"Discrimination and Perceptual Knowledge"

In this paper Goldman is explicitly concerned with modifying the causal theory of knowledge to deal with counterexamples based on external defeaters.¹⁵ He intends to deal with external defeaters by developing an account of perceptual knowledge using the concept of a "relevant perceptual alternative". The general idea is that for S to know p , S must be able to discriminate situations in which p is true from perceptually equivalent states of affairs in which p is false. A first formulation of this idea is as follows:

X knows p only if the actual state of affairs in which p is the case is discriminable by X from every relevant possible alternative in which p is not the case.¹⁶

¹⁵ Goldman 1976: 772-3. Here he considers the "fake-barns" case and writes that his causal analysis cannot handle it.

¹⁶ *Ibid.*: 774. Goldman actually writes "*a* relevant possible alternative", as opposed to "*every*". This is probably a typographical mistake, since the analysis he goes on to develop clearly requires the universal quantifier.

Goldman's reliabilism

Goldman sees the relevant alternatives account of knowledge as closely connected to the account of knowledge in terms of reliability. As he writes:

To be reliable, a cognitive mechanism must enable a person to *discriminate or differentiate* between incompatible states of affairs. It must operate in such a way that incompatible states of the world would generate different cognitive responses. Perceptual mechanisms illustrate this clearly. A perceptual mechanism is reliable to the extent that contrary features of the environment (e.g., an object's being red, versus its being yellow) would produce contrary perceptual states of the organism, which would, in turn, produce suitably different beliefs about the environment.¹⁷

This suggests strongly that Goldman takes it that a process is reliable *just in case* it is able to discriminate between incompatible states of affairs.¹⁸ We should note that an account of knowledge in terms of reliability has the potential to provide an explanation of a priori knowledge, for, since a reliable connection does not in general have to be a causal connection, there seems to be no conceptual problem with the idea of being reliable about the objects of a priori knowledge.¹⁹

The initial formulation of Goldman's idea needs cashing out with an account of discrimination, and with an account of relevance. We will deal with these in turn.

Discrimination

On Goldman's account, external defeaters undercut knowledge because the subject cannot discriminate the actual situation from a possible alternative situation in which their belief is false. If subject X cannot discriminate (by perception alone) a state of affairs ϕ from state of affairs κ , then ϕ and κ are *perceptually equivalent* for X. The concept of perceptual equivalence is to be explained using the concept of a *perceptual state of affairs*: on Goldman's account this is an ordered triple, consisting of

¹⁷ *Ibid.*: 771.

¹⁸ Modulo, that is, considerations concerning relevance to be introduced shortly.

¹⁹ Casullo 1988: 211.

Goldman's reliabilism

(i) an object; (ii) the set of all properties of that object (some but not all of which need to be perceptually accessible properties); and (iii) a specification of the environmental features which have a bearing on the way the object appears to the subject - he calls this a DOE relation (Direction-Orientation-Environment relation).²⁰

Using the notion of a perceptual state of affairs, a first suggestion at the relationship between perceptual equivalence and discriminability can be stated:

(PE1) For subject X , objects b and c , maximal set of properties J and K , and DOE relations R and R^* : $\langle b, J, R \rangle$ is a perceptual equivalent, for X , to $\langle c, K, R^* \rangle$ iff X cannot discriminate between $\langle b, J, R \rangle$ and $\langle c, K, R^* \rangle$.

As Goldman points out, this definition of a perceptual state of affairs is strictly speaking inadequate: a full specification will need to take account of the state of functioning of the subject's perceptual organs, and the subject's powers of attentive discrimination. A full discussion of these factors can be passed over, but one point is particularly important: while (PE1) does not require that $\langle b, J, R \rangle$ and $\langle c, K, R^* \rangle$ be *qualitatively identical* perceptual states of affairs, the similarity between them cannot be weakened too much, or intuitively non-equivalent perceptual states of affairs might be classed as equivalent.²¹ We need a stable account of perceptual equivalence which stops short of qualitative identity. Goldman's proposed solution to this is that qualitative identity of perceptual states of affairs can be dropped so long as no features relevant to the formation of the belief in question are thereby rendered discriminable. The full definition of perceptual equivalence then becomes:

(PE2) If object b has the maximal set of nonrelational properties J and is in DOE relation R to subject X at time t , if X has some percept P at t that is perceptually caused by b 's having J and being in R to X at t , and if P noninferentially causes X to believe (or sustains X in believing) of object b that it has property F , then

$\langle c, K, R^* \rangle$ is a *perceptual equivalent* of $\langle b, J, R \rangle$ for X at t relative to property F iff

(i) if at t object c had K and were in R^* to X , then this would perceptually cause X to have some percept P^* at t ,

²⁰ Goldman 1976: 780-1.

²¹ *Ibid.*: 782.

Goldman's reliabilism

- (ii) P^* would noninferentially perceptually cause X to believe (or sustain X in believing) of object c that it has F , and
- (iii) P^* would not differ from P in any respect that is causally relevant to X 's F -belief.²²

In moving from (PE1) to (PE2) Goldman avoids a problem which he has apparently not recognised. As we saw above, Goldman wants to give a reductive definition of knowledge; to be successful, then, the definition arrived at must not appeal to any epistemic concepts. DeRose, however, has suggested that an appeal to "discrimination" masks a tacit appeal to the concept of knowledge.²³ Applied to present case, DeRose's point is that to say that S cannot discriminate state of affairs ϕ from state of affairs κ is to say no more than that in κ , S does not *know* she is not in ϕ . If this is right, Goldman is barred from appealing to (PE1); however, as DeRose goes on to point out, Goldman's second formulation of perceptual equivalence avoids appeal to discrimination by appealing to what the subject would believe in counterfactual situations where her sensory stimulation was different.

The analysis of perceptual equivalence is used in giving a general analysis of perceptual knowledge.

At t X noninferentially knows of object b that it has property F iff

- (1) for some maximal set of nonrelational properties J and some DOE relation R , object b has (all the members of) J at t and is in R to X at t ,
- (2) F belongs to J ,
- (3) (A) b 's having J and being in R to X perceptually causes X at t to have some percept P ,
(B) P noninferentially perceptually causes X at t to believe (or sustains X in believing) of object b that it has property F , and
(C) there is no alternative state of affairs $\langle c, K, R^* \rangle$ such that (i) $\langle c, K, R^* \rangle$ is a perceptual equivalent of $\langle b, J, R \rangle$ for X at t relative to property F , and (ii) F does not belong to K .
- (4) X 's propensity to form an F -belief has an appropriate genesis.²⁴

²² *Ibid.*: 783.

²³ DeRose 1995: 17 n.23.

²⁴ Goldman 1976: 785, 789.

Goldman's reliabilism

The explicit appeal to "relevance" has been dropped, but is implicit in clause 3(C): here, the appeal to an "alternative state of affairs" is clearly meant to import an appeal to *relevant* alternative states of affairs. On this definition, clause 2 requires that that which is known be true; clause 3(B) requires that that which is known be believed. If we abstract from the truth clause we are left with Goldman's 1976 account of warrant. A belief is warranted, on this account, just in case: (i) it is caused by the fact which the belief is about (clause 3(A) and 3(B)); (ii) there are no relevant perceptually equivalent defeaters (clause 3(C)); and (iii) the propensity to form the belief is appropriately brought about (clause 4).

The discussion of Goldman's account of perceptual knowledge needs completing by a discussion of relevance, and by an explanation of clause 4.

Relevance

It is crucial to Goldman's account that he offer a satisfying account of relevance. If the set of alternative situations is left unrestricted, it will include situations in which the subject is a brain in a vat, being perfectly deceived through computer-driven simulation, into believing that she is not a brain in a vat but living a perfectly normal life. Brain-in-a-vat style sceptical scenarios are defined as being perceptually equivalent to the actual situation: if the set of alternative situations is left unrestricted then clause (3C) can never be filled, and Goldman's definition of perceptual knowledge will issue in scepticism. The tuning of the "relevance" relation is therefore crucial to an account of perceptual knowledge which appeals to relevant alternatives. If the relevance relation is set too weak, and certain intuitively non-relevant possible situations are treated as relevant, the account will undergenerate (that is: it will not test positive for knowledge in cases which intuitively are cases of knowledge). If the relevance relation is set too strong, and certain intuitively relevant possible situations are treated as non-relevant, then the account will overgenerate (that is: it will test positive for knowledge in cases which intuitively are not cases of knowledge).

Goldman's reliabilism

Further, it is crucial for Goldman that relevance turn out to be naturalistically acceptable, that is, definable in non-epistemic, non-normative terms. Goldman suggests that the relevant alternatives are determined by the context - physical, linguistic and psychological - in which the subject is embedded.²⁵ Though Goldman does not give a fully worked out theory of relevance, he does suggest that certain "regularities" govern what is to be counted as relevant. We can summarise his suggestions as follows. For circumstances ϕ and κ :

- if, given κ obtains, it is *likely* ϕ obtains, then ϕ is relevant to κ ;
- if ϕ is similar to κ , then ϕ is relevant to κ ;
- if S believes ϕ is relevant, then ϕ is relevant.²⁶

Goldman's account of relevance, however, is left deliberately vague and unfinished; Goldman does not commit himself to these suggestions, nor does he suggest that they are exhaustive. We shall return to the problem of relevance in the next chapter.

Clause 4

Clause 4 is needed to rule out counterexamples such as the following:

Suppose Sam's "schemata" [an hypothetical iconic set of features used for perceptual recognition] of Judy and Trudy have hitherto been indistinct, so Judy-caused percepts sometimes elicit Judy-beliefs and sometimes Trudy-beliefs, and similarly for Trudy-caused percepts. Today Sam falls down and hits his head. As a consequence a new feature is "added" to his Judy-schemata, a mole-associated feature. From now on he will believe someone to be Judy only if he has the sort of percept that would be caused by a Judy-like person with a mole over the left eye. Sam is unaware that this change has taken place and will remain unaware of it, since he isn't conscious of the cues he uses. Until today, neither Judy nor Trudy has had a left-eyebrow mole; but today Judy happens to develop such a mole. Thus, from now on Sam can discriminate Judy from Trudy.²⁷

²⁵ Goldman 1976: 776.

²⁶ *Ibid.*

²⁷ *Ibid.*: 789.

Goldman's reliabilism

Although Sam can now correctly identify Judy, intuitively, his first identification of her at least will not count as knowledge. The structure of this case is as follows: Sam has the ability to discriminate states of affairs in which the proposition he believes is true, from relevant counterfactual states of affairs in which it is false (that is, Sam meets the requirements of Goldman's 1976 analysis, bar clause 4); however, this ability has been brought about serendipitously, and therefore beliefs produced as a result of it do not count as knowledge. Let us call the problem of how to distinguish appropriate from inappropriate methods of belief-production the problem of *deviant causal chains*. As Goldman admits, his treatment of the problem of deviance by the introduction of clause four does no more than identify the problem. The nature of the problem is clarified, and responses proposed, in his later work.

In conclusion to this discussion of Goldman's 1976 account of perceptual knowledge, we should note the following.

First: pending further specification of "relevance" and "deviance", Goldman's analysis deals with the problem of counterexamples based around external defeaters. In case G, the possibility that the barn S sees is a fake must be considered a relevant perceptual alternative which she cannot discriminate from actuality. Therefore (3C) is violated.²⁸ Note that since Goldman ties the reliability of a process to its ability to discriminate, it seems that he must say that S's visual belief forming process is *not* reliable in case G.

There is no explicit mention of the problem of internal defeaters, though Goldman's third suggestion as regards relevance suggests an interesting line of response: perhaps an analysis could be developed on which S's *belief* that a defeating circumstance obtains creates that circumstance as an unexcludable relevant alternative, and so defeats S's knowledge. However, this suggestion will have to be developed before it can protect the 1976 account from counterexamples based on internal defeaters. In case F, S does not have knowledge, but there seems to be no relevant indiscriminable situation in which S's belief is false - at least, on

²⁸ *Ibid.*: 786-7.

Goldman's reliabilism

the account of relevance that Goldman provides there seems to be no grounds for saying that such a situation is relevant. Case F is a case of memory which Goldman's 1976 theory is not designed to address, but we could clearly develop perceptual cases that would be structurally equivalent. Intuitively, we are inclined to say that the reason S does not know in case F is because her belief is unjustified relative to her total available information (which is, as it happens, misleading). We should therefore hold out hope that the account of justification in "What is Justified Belief?" will address cases such as F.

Second: the 1976 account of knowledge secures the infallibility of warrant through two independent routes. First, it incorporates the feature, central to the earlier causal theory, that a belief is known only if it is causally connected with the fact which it concerns. We have already argued that this generates an infallibilist account of warrant. Further, though, by requiring the absence of relevant counterfactual defeating situations, Goldman secures infallibility by a second route, as we will now argue.

Clause (3C) in Goldman's definition of warrant is a requirement that there be no relevant perceptually equivalent situation in which the perceptual belief in question be false. That is, Goldman requires that, where S believes p , if there are any relevant perceptually equivalent situations in which p is false but S believes p anyway, then S's belief in p is not warranted. Contraposing: if S's belief in p is warranted then there is no relevant perceptually equivalent situation where p is false but S believes p anyway. We can express this latter proposition as follows (*key*: variables range over worlds; Fw = S's belief in p is warranted in w ; $G(v,w)$ = world v is perceptually equivalent to and relevant to world w ; Hv = p is true in v):

$$\forall w (Fw \rightarrow \neg \exists v (G(v,w) \& \neg Hv))$$

Given the plausible claim that each world is relevant to and perceptually indiscriminable from itself, that is:

$$\forall w G(w,w)$$

it follows that:

$$\forall w(Fw \rightarrow Hw)$$

that is: in all worlds, if a belief is warranted then it is true. Since the necessity operator is equivalent to universal quantification over (accessible) worlds, this conclusion is just the claim that: necessarily, if S's belief is warranted then S's belief is true. But we have said that a property of beliefs is infallible just in case, necessarily, it only attaches to true beliefs. Hence, again, on Goldman's account warrant is infallible.

Third: this account of knowledge is concerned only with perceptual knowledge. The account therefore stands in need of generalisation to other classes of knowledge such as knowledge from memory, inference and a priori knowledge. We will return to this after discussion of the account of justification.

"What is Justified Belief?"

As we have already noted, the reason why S fails to have knowledge in case F is most naturally described in normative terms. S, we are inclined to say, is simply wrong to ignore her parents' misleading testimony. She *ought* to take it into account, even though in ignoring it she puts herself in what is, objectively speaking, a stronger epistemic position with respect to her beliefs about the past. S's belief is not an item of knowledge because it is, in the intuitive sense, not justified; and we have argued that we can admit this without committing ourselves to the claim that justification is in general necessary for knowledge.²⁹

Goldman, however, cannot appeal to the notion of normative justification unreduced: such an appeal would contravene his naturalist principles. Accordingly

²⁹ See chapter six.

Goldman's reliabilism

he seeks to give an analysis of the concept of justification in nonepistemic terms. For Goldman, a certain sort of naturalistically acceptable justification *is* necessary for knowledge, and he tries to explain this notion in terms of the concept of reliability. Reliable belief forming processes, he claims, are such that they confer justification on the beliefs that they produce. The key question is whether the notion of justification which Goldman develops can be used to explain internal defeat.

Goldman understands processes as *functions*, generating a mapping from certain input states onto certain output states.³⁰ The possible inputs to the processes which interest us are beliefs or experiences; the outputs are beliefs. Reliability is a property of a process type. A process type is reliable iff it has the statistical property of producing true beliefs to a suitably high degree (just *what* degree we require is not stated). As Goldman points out, we need to be able to talk of processes *causing* beliefs; but processes are construed as types, and types are abstract objects which presumably cannot enter into causal relations: so there is some awkwardness in this. Goldman suggests that the problem is really just one of sloppy expression.

...when we say that a belief is caused by a given process, understood as a functional procedure, we may interpret this to mean that it is caused by the particular *inputs* to the process (and by the intervening events 'through which' the functional procedure carried the inputs into the output) on the occasion in question.³¹

Some terminology is required before we can state Goldman's definition. A *belief-dependent* process is a belief forming process that takes beliefs as at least part of its input. A *belief-independent* process is one that does not take any beliefs as its input. Since the fact that a belief-dependent process produces a false belief when some of its input beliefs are false does not count against the reliability of that process, we also need a concept of conditional reliability: a process is *conditionally*

³⁰ Goldman 1992: 116. ("What is Justified Belief?" was originally published in 1979; page references are to the reprint in Kornblith (ed.) 1992.)

³¹ *Ibid.*: 117.

Goldman's reliabilism

reliable iff it is reliable given that its input beliefs are true. Using these definitions, Goldman's initial recursive definition of "S's belief in p at t is justified" is as follows:

(WJB1) If S's belief in p at t results immediately from a belief-independent process that is unconditionally reliable, then S's belief in p at t is justified.

(WJB2) If S's belief in p at t results immediately from a belief-dependent process that is at least conditionally reliable, and if the beliefs (if any) on which this process operates in producing S's belief in p at t are themselves justified, then S's belief in p at t is justified.³²

This account, however, simply does not touch the problem of internal defeaters. In case F S's belief in p meets the requirement of this definition but we do not want to say that it is justified. Goldman recognises this and proposes a modification to his account. The reason why S's belief in case F is unjustified, he suggests, is because she has strong evidence that she doesn't, but should have, used.³³ So stated, this diagnosis clearly turns on normative considerations, but Goldman takes it to be equivalent to the following non-normative claim: there is a belief forming process available to her (the process of forming beliefs based on listening to her parents) which S doesn't use, but which would have given importantly different results had she used it. With this in mind, Goldman reformulates the base clause principle to address the problem of internal defeat.

(WJB1*) If S's belief in p at t results from a reliable belief forming process, and there is no reliable or conditionally reliable process available to S which, had it been used by S in addition to the process actually used, would have resulted in S's not believing in p at t , then S's belief in p at t is justified.

There are a number of problems with Goldman's account of justification. First, this reformulated account is still susceptible to counterexample, as the following BonJour-style example shows.³⁴

³² *Ibid.*: 119.

³³ *Ibid.*: 126.

³⁴ From BonJour 1985.

Goldman's reliabilism

H S reads p in a respected newspaper and so comes to believe p . In fact p is false, this being one of the newspaper's rare errors. On the table at which S is reading, there happens to stand an infallible crystal ball, which, if S had consulted it, would have revealed the truth about p . S has heard stories about the powers of the crystal ball but does not believe them and so does not consult the ball.

Goldman's revised condition for justification is not filled, since the process of consulting the ball is available to S. So Goldman's analysis has it that S is not justified in believing p on the basis of the newspaper; but, intuitively, S is justified in believing in p , since (we may assume) S is justified in discounting fantastical stories about the powers of the ball. So Goldman's account of internal defeat is inadequate: he fails to give a reductive account of the concept of justification.

Further, by shifting away from the "caused by the fact" requirement in the two earlier accounts, Goldman loses the feature of infallibility from his account of justification. Since, roughly speaking, justification is conferred on a belief when that belief is caused by a reliable process, on the rare occasion that a reliable process produces a false belief (as the newspaper did in case H), that belief is justified.³⁵ Since justification can attach to false beliefs it is hard to see how the addition of a clause requiring justification to a basic reliabilist account of warrant can do any work in securing infallibility.

The transition to a reliable process account brings with it a further problem, considered by some to be insurmountable.³⁶ This is the problem of how to individuate belief forming processes with respect to their reliability. We have said that a process is reliable just in case the beliefs it produces are metaphysically likely to be true. But any given belief can be considered to be the outcome of any number of processes, and these processes will vary widely in the proportion of true beliefs that they produce, and so vary widely in their degree of reliability. Since we are linking justification to reliable production, whether a belief is justified will depend crucially on which process we take it to be produced by.

³⁵ Goldman 1992: 116.

³⁶ Pollock 1986 argues against reliabilism on the basis of the generality problem. I will shortly argue that Pollock's objections are not convincing.

Goldman's reliabilism

This is known as the *generality problem*. An illustration may help. Say S spots a pink elephant and so comes to believe that there is a pink elephant before her. Which process produced S's belief? We may imagine that S's visual acuity is high, and that the process of S forming beliefs based on what appears to her visually is extremely reliable: if we take S's belief to be caused by this process, S will be justified in believing that there is a pink elephant before her. But we could also assign S's belief to the process of people in general coming to believe that there is a pink elephant before them after *seeming* to see a pink elephant before them. This process, we may suppose, is associated primarily with alcohol detoxification and is highly unreliable; so if S's belief is a product of this process then it is not justified. S's token belief that there is a pink elephant before her can be assigned to either of these processes, and infinitely more. We currently lack a criterion with which to decide to which process S's belief should be assigned. This problem will be returned to in the context of discussion of the account in *Epistemology and Cognition*.

Let us sum up the conclusions drawn from the discussion of Goldman's earlier accounts of knowledge and justification, in order to identify what problems the account in *Epistemology and Cognition* needs to address if it is to be successful. We need an epistemology that

- secures the infallibility of warrant
- yields a definition of knowledge which excludes counterexamples based on external defeaters and internal defeaters
- solves the generality problem
- explains the difference between deviant and nondeviant belief forming processes
- leaves room for a priori knowledge.

The account of discrimination and perceptual knowledge has the potential to address the first two problems via an appeal to relevant counterfactual circumstances: indeed, given that the paper on justification gives a very unconvincing account of internal defeat, it might seem that the most promising route available to us would be to concentrate on developing the former account. But an account of justification is also important. As we have formulated it, the

Goldman's reliabilism

problem of nondeviance concerns drawing a distinction between different sorts of processes; the problem is therefore best addressed within an epistemological framework which assigns a role to processes. The idiom of processes will also be crucial for the account of experience-independence in terms of nonexperientiality which we will discuss in chapter nine. However, the introduction of belief forming processes requires us to address the generality problem. We turn now to see how Goldman's most detailed account of knowledge addresses these problems.

Epistemology and Cognition

Goldman does not commit himself to a full analysis of knowledge in *Epistemology and Cognition*; he gives only a series of necessary conditions. However, he does not argue that an analysis of knowledge cannot in general be given³⁷, and, indeed, he even suggests that the possibility of a naturalistic response to certain forms of scepticism *requires* that an analysis be given.³⁸ Therefore, I will take the step of assuming that the necessary conditions Goldman offers are intended to be jointly sufficient. This assumption may be illegitimate: it may be unfair to press Goldman's conditions into the demanding role of an analysis. However, if we are unable so to interpret them then the key question of whether warrant, on Goldman's view, is infallible, is simply ducked without explanation. So I think we should take Goldman to be proffering an analysis; by focusing on whether reliabilism can meet the infallibility constraint we will be able to arrive at a clear view of the serious problems facing the reliabilist account of knowledge in general.

Knowledge

Goldman's basic necessary condition for knowledge is, on the surface, much simplified from the 1976 account. It is:

³⁷ Cf. Williamson 1995 and Botterill and Carruthers 1999, who argue that only necessary conditions can be given a priori.

³⁸ Goldman 1986: 36.

Goldman's reliabilism

(EC1) S knows p only if S's belief in p results from a reliable belief forming process.³⁹

But, as Goldman points out, this needs filling out by a detailed account of reliability itself. Goldman argues that the reliability of a belief forming process must be understood as a *propensity* of that process to form true beliefs; equivalently, the reliability of a process type α is given by the proportion of true beliefs produced by processes of type α in possible situations.⁴⁰ He distinguishes between two grades of reliability. A process is *globally* reliable just in case it produces the required proportion of true beliefs in an unrestricted class of possible situations; a process is *locally* reliable just in case it produces the required proportion of true beliefs in the class of relevant possible situations.⁴¹ Goldman requires both local and global reliability of the processes which are apt to produce knowledge, though he places emphasis on the requirement of local reliability. Schematically, local reliability can be represented as follows:

process α is locally reliable iff in *most* relevant situations: if α produces a belief in p then p is true.

The requirement of local reliability, he claims, is drawn from and equivalent to the 1976 account of perceptual knowledge. However, this does not seem to be quite right: there seems to be a significant difference between them. For the 1976 account embodied an infallibilist account of warrant. On that account, if S knows p then there are *no* relevant perceptually equivalent counterfactual situations in which S falsely believes p . Schematically, the definition of local reliability arising from the 1976 account is as follows:

process α is locally reliable(1976) iff in *all* relevant situations: if α produces a belief in p then p is true.

³⁹ Goldman 1986: 52.

⁴⁰ The reliability of a process cannot be calculated in terms of the process's performance in the *actual* world, since a case of bad epistemic luck could result in an intuitively reliable process only actually producing false beliefs. Cf. Goldman 1986: 49.

⁴¹ *Ibid.*: 45.

The 1976 and 1986 accounts therefore differ over whether a locally reliable process can produce a false belief. To emphasise this difference, consider the following quotation.

For a belief to count as knowledge, I am arguing, it must be caused by a reliable process. Exactly *how* reliable I have not said. Nor do I think this can be answered with precision. The knowledge concept is vague on this dimension, and an analysis need not impose more precision than the common sense concept contains.⁴²

To say that a process must be infallible if the beliefs it produces are to count as knowledge is to give a precise answer to this question: so, Goldman seems to be claiming not to be committed to the view that belief forming processes are infallible. On the 1986 account, then, infallibility is not built directly into the requirement of local reliability. If the account of warrant we are attributing to Goldman is to be infallibilist, there must be some other mechanism in his account which has the consequence that when a locally reliable process produces a false belief, that belief does not get warranted. We shall survey the rest of Goldman's 1986 account of knowledge to see if this condition is met. The conclusion to be drawn is that it is not met; however, it could be met by bringing his newer account more in line with the suggestions offered in the 1976 paper. In the next chapter we will focus on how Goldman's account of knowledge could be augmented with a worked-out theory of relevance, and we will ask whether the resulting account is properly naturalistic.

Goldman supplements the necessary condition for knowledge given in (EC1) in three ways. The first of these is a proposed solution to the generality problem. The second is a requirement that the belief forming process be nondeviant, and he sketches a solution to the problem of deviance. The third is the requirement that the belief produced be justified, and he offers an updated reliabilist account of justification. We will discuss these in turn.

⁴² Goldman 1986: 51.

Goldman's reliabilism

The generality problem

The generality problem has been taken to be insuperable for reliabilists. Discussing a reliabilist account of justification (essentially that of Goldman 1992) Pollock argues that in individuating a belief forming process we have to take *all* information into account; as a consequence, each process type will have only a single instance.⁴³ Moreover, included in that individuating information will be facts about the truth value of the belief produced: hence the reliability of the single instance of the process expressed as a fraction, will be either 1, if the belief produced is true, or 0, if the belief produced is false. Hence a process is absolutely reliable if it produces a true belief; and hence, only processes which are absolutely reliable are able to confer justification on the beliefs they produce. Let us call this the *single case* problem.⁴⁴

It is not easy to determine what Pollock's objection is meant to come to. He writes that what is objectionable about this conclusion is that it makes justification infallible:

This reliabilist criterion entails the absurd consequence that in order for a belief to be justified it must be true.⁴⁵

The intuitive notion of justification indeed does not require that justified beliefs be true; hence Pollock's point would be significant if it could be used to show that reliabilists are committed to the infallibility of the intuitive notion of justification. But as we have argued, we cannot take Goldman to be advancing an account of the intuitive notion, for that notion is not necessary for knowledge, whereas Goldman seeks to give an account of justification such that justification is necessary for knowledge. Given that Goldman seems to be using "justification" in a technical sense, it is not clear whether Pollock's observation counts as an objection to it. Certainly, if by "justification" Goldman means something like our "warrant", Pollock's point, if correct, would seem to be a strike in its favour.

⁴³ Pollock 1986: 118.

⁴⁴ The terminology is from Goldman 1986.

⁴⁵ Pollock 1986: 118.

Goldman's reliabilism

Goldman himself gives a more convincing account of the problem posed by the single case problem. The problem as he sees it is this: if we individuate processes so narrowly that any true belief arrived at counts as reliably produced, we will bring ourselves into conflict with intuition.⁴⁶ So, *guessing* is, let us say, not a reliable belief forming process; hence true beliefs formed by guessing are not known. But on the maximally precise individuation of belief forming processes envisaged by Pollock, a token process of guessing truly will indeed count as reliable, and yet it clearly is not. The problem for reliabilists, then, is to block the requirement that the maximum amount of information be taken into account in individuating processes.

Goldman suggests that we have already avoided the single case problem by taking reliability to be a propensity for tokens of a process type to produce true beliefs in counterfactual situations.

Now the Single Case problem arises only if global reliability is determined exclusively by actual frequencies. As suggested... however, a propensity approach is preferable.⁴⁷

What he seems to mean is that the definition of global reliability in terms of counterfactual situations rules out the option of individuating types so narrowly that they have but one instance. However, he gives no defence of this, and without a defence his response might seem implausible: after all, a really specific individuation of a belief forming process type would also specify which single world that globally reliable process type was instantiated in. Then it will be trivially counterfactually reliable, relative to that world, given that it issues in a true belief in that world.

However, we may be able to interpret Goldman more favourably; perhaps he has in mind something on the following lines. The fundamental idea of reliabilism is that reliable *causation* of a belief is a necessary condition for knowledge. On an influential theory, causation is analysed counterfactually⁴⁸:

⁴⁶ Goldman 1986: 50.

⁴⁷ *Ibid.*

⁴⁸ Lewis 1973.

contra Hume, it is claimed that if we cannot talk of the relationship between cause c and effect e in counterfactual situations then we cannot capture the causal relationship we ascribe to them in actuality.⁴⁹ Perhaps, then, Goldman's idea is that if we individuate belief forming processes so narrowly that each type has only one instance, we will be unable to draw a distinction between the belief so produced being produced by accident, as opposed to being caused. We lack any counterfactual situations to take as a point of comparison, and this is simply to give up the idea that the belief was caused. Hence the idea of a causal theory of knowing requires that we individuate processes in a way that allows a degree of re-identification of the process in counterfactual situations. So, the single case problem does not arise.

But this still leaves us with the problem of how we should individuate processes. Goldman suggests, plausibly enough, that we should individuate process types quite narrowly, in order that we can draw the required distinction between processes that intuitively do yield knowledge and those which intuitively do not.⁵⁰ Goldman's suggestion is as follows: the process type whose reliability must be taken into account in assessing a token belief for positive epistemic status is *the narrowest process type that was causally operative in producing the belief token in question*.⁵¹ Goldman illustrates his proposal with an example of an hypothetical mechanism for forming perceptual beliefs. We are to imagine that this mechanism takes sensory information about some object O as input and matches it to various templates, each representing some category C . There is a value T (where $0 < T \leq 1$) such that if the input matches the template of C to a degree T or more, the mechanism outputs the belief that the object O belongs to category C . If T is very low, then the mechanism will output the belief that O belongs to C even under

⁴⁹ The Humean account is reductionist: c caused e just in case there is constant conjunction of c and e in actuality, with appropriate temporal priority, and we have come to expect es to follow cs (the metaphysical aspect of causation, the modal relationship between cause and effect, is substituted for by a psychological property, namely expectation).

⁵⁰ Goldman 1986: 50.

⁵¹ *Ibid.* It is reasonable to assume that Goldman has in mind a criterion for narrowness as follows: belief forming process type α is narrower than type β just in case α has fewer instances than β . Though I think this is intuitively clear, strictly speaking α and β will have infinite instances. To capture the intuitive clarity of the criterion, it is best to shift to an illustration in terms of possible worlds: α is a narrower type than β just in case the sphere of worlds in which α is instantiated is smaller than the sphere of worlds in which β is instantiated.

Goldman's reliabilism

highly adverse visual conditions, and so even if the mechanism outputs a true belief that O belongs to C on some specific occasion, the general unreliability of the mechanism will stop us assigning that true belief the status of knowledge. But we can also imagine that on some occasion perceptual conditions are good, and the stimulus from O matches C to a very high degree - say .99. Under these circumstances, if O does belong to C, and the mechanism outputs an appropriate belief, we will want to say that the belief is known: but if the process type selected still includes the very low value of T, the mechanism will not be deemed adequately reliable. Goldman's suggestion then operates as follows: in the case where the degree of match is .99, the narrowest process type that is causally operative in producing the belief is the property of producing-a-belief-when-the-degree-of-match-is-.99. Although the belief can also be taken to be a product of the wider process type of producing-a-belief-when-the-degree-of-match-is-T, this process, since it is wider, is deemed causally inoperative. Thus Goldman's suggestion allows us a principled way of selecting processes, which fits our intuitions about whether the belief is known.⁵²

Goldman's suggestion is not arbitrary. As already suggested in connection with Kitcher's definition of a priori knowledge, there is an intuitive distinction between the cause of an event, the background conditions against which the cause operates, and other conditions irrelevant to the causal process. Belief forming processes are of epistemic interest only in so far as they cause beliefs; since it is their causal role that is crucial, we need not concern ourselves with more information about them than is relevant to that role. Pollock's charge that we must take the maximum amount of information into account in individuating processes would have us count as significant, information which does not concern the process considered as a cause. On Goldman's proposal, we count as significant only information about the process considered as a cause of belief, and this respects the intuitive distinction between cause, background condition, and irrelevant condition.

⁵² Goldman 1986: 50-1.

Deviant processes

The 1986 account provides a more worked out account of deviant processes. Goldman discusses two ways in which a reliable belief forming process may be unable to warrant the beliefs it produces.⁵³

First, where a belief forming process takes as input other beliefs, that process, even if locally and globally reliable, will not confer warrant on the beliefs it produces unless the input beliefs themselves were warranted. Thus, as Goldman puts it, whether a given belief is warranted depends on the remote cognitive ancestry of that belief.

Second, no process, even if locally and globally reliable, will confer warrant on the beliefs it produces if the process itself was deviantly produced. Let us call a process which produces belief forming processes a *second order belief forming process*. Goldman now requires that a belief forming process α confers positive epistemic status on the beliefs it produces only if α was brought about by an appropriate second order process. The following case illustrates:

- J S learns an algorithm A from a wholly unreliable quack mathematician, Elmer Fraud. The vast majority of algorithms taught by Fraud are defective; however, A is a rare correct one, and S learns it and uses it to form a true belief p .⁵⁴

Intuitively, S does not know p . This is because the second order belief forming process - that of forming belief forming processes based on instruction by Fraud - is unreliable. Goldman accordingly proposes an addition to the original condition for knowledge given in (EC1):

- (EC2) An acquired belief forming process can generate knowledge only if it is acquired or sustained by an appropriate second order process.⁵⁵

⁵³ *Ibid.*: 51.

⁵⁴ *Ibid.*: 51-2.

⁵⁵ *Ibid.*: 52. I pass over Goldman's distinction between belief forming processes and belief forming methods as a detail we do not need to consider here.

Goldman's reliabilism

What is crucial, of course, is the explanation of "appropriateness". Goldman makes two suggestions, but does not adjudicate between them. First: a second order belief forming process is reliable just in case the proportion of reliable processes it produces is very high. Second: a second order belief forming process is reliable just in case it only generates processes which are more reliable than the ones previously in use in the same context.⁵⁶ Note that either of these suggestions will solve the Judy/Trudy case discussed in the long quotation from Goldman 1976.⁵⁷ The second order process of forming belief forming processes through being bumped on the head is reliable in neither of the senses given here.

The requirement of appropriate higher order justification suggests a regress: we might think that a higher order process that meets Goldman's stated criteria for appropriateness would be unable to confer warrant-conferring-status on the belief forming processes it produces if it itself was unreliably produced. However, this regress does not have to be infinite or vicious. Goldman suggests that as we ascend the hierarchy, we come quickly to higher order belief forming processes which are deep-seated cognitive mechanisms whose appropriateness, or lack of, is something to be determined by empirical science. These deep-seated processes anchor the positive status of the processes they produce. Goldman leaves open the possibility that investigation will reveal that the higher order processes are not appropriate, and that in fact none of our beliefs has warrant. That this possibility stands is a consequence of Goldman's naturalism. A successful naturalistic analysis of knowledge, he claims, can only be expected to show that it is *possible* that we do have knowledge; a demonstration that we actually have it - that is, a refutation of a certain form of scepticism - is not something that can be carried out a priori.⁵⁸

Both the conditions for nondeviance that Goldman offers have the consequence that we do not automatically know what we know. It will often not be possible for us to determine the remote cognitive pedigree of the beliefs our

⁵⁶ *Ibid.*: 53.

⁵⁷ See page 150.

⁵⁸ Goldman 1986: 55-57 and ch.2.

processes operate on; in which case, again, we will be unable to determine a priori whether a given belief which we entertain constitutes a case of knowledge.⁵⁹

Justification

Like the earlier account, Goldman's 1986 account of justification is based on reliability, and justification is intended to be necessary for knowledge.⁶⁰ The new account differs from the old in being developed within a framework of rules ("justification rules" or "J-Rules") and in being augmented by a "no-undermining" clause. The rule framework is introduced to capture the evaluative aspect of justification. In saying that a subject's belief is justified, Goldman allows, we say that the subject is in some sense right to entertain it; however, the concept of justification embodied in the rule framework is not meant to be regulative or action guiding.⁶¹

A person need not even understand the rules, and if he does, he need not be able to apply them in the process of belief formation.⁶²

Rather, these rules constitute principles for third-person evaluation of beliefs; they are standards against which an appraiser can measure a subject's beliefs. Justification, in Goldman's sense of the term, is not action guiding: it is evaluative but nonregulative.

Goldman's initial definition of justification is as follows:

- (P1) S's believing p at t is justified if and only if: S's believing p is permitted by a right system of J-rules.⁶³

⁵⁹ Since Goldman at the same time maintains that we do have knowledge, he is committed to denying the S4 axiom for knowledge (that is, the *KK thesis*: $Kp \rightarrow KKp$). This puts him squarely in the epistemic externalist camp.

⁶⁰ *Ibid.*: 58.

⁶¹ *Ibid.*: 25.

⁶² *Ibid.*: 59.

⁶³ *Ibid.*

Goldman's reliabilism

The key question, of course, is what counts as a right system. After an extended argument by elimination, Goldman opts for the reliabilist criterion. His final formulation of the criterion is as follows:

(ARI) A J-rule system R is right if and only if R permits certain (basic) psychological processes, and the instantiation of these processes would result in a truth ratio of beliefs that meets some specified high threshold (greater than .50).⁶⁴

This, in essence, is unchanged from the earlier account, and is still vulnerable to objections related to case F, as Goldman recognises. In case F, we have suggested, it is most natural to blame S's lack of knowledge on normative failings. To handle cases such as F, Goldman introduces the "no-undermining" clause. The completed definition of justification is then as follows:

(P3) S's believing p at t is justified if and only if:
(a) S's believing p is permitted by a right system of J-rules, and
(b) this permission is not undermined by S's cognitive state at t .⁶⁵

It is crucial, of course, that Goldman cash out the no-undermining clause in non-normative terms. He does this by what seems to be intended as an enumeration of cases. The three types of undermining he discusses are:

- S's justification for believing p is undermined if S permissibly believes that her belief in p is not permitted.
- S's justification for believing p is undermined if S impermissibly believes that her belief in p is not permitted.
- S's justification for believing p is undermined if S believes that conditions C are not satisfied, where in fact the satisfaction of C is a necessary condition for S's belief in p to be justified.⁶⁶

This clause handles case F, since in this case the parental testimony permits S to believe that her belief about her childhood is not permitted. However, Goldman's

⁶⁴ *Ibid.*: 106.

⁶⁵ *Ibid.*: 63.

⁶⁶ *Ibid.*: 62.

Goldman's reliabilism

no-undermining clause has an *ad hoc* air: it is not clear that Goldman has identified a general strategy for dealing with cases of defeat through what we would ordinarily describe as normative failings. Cashed out, Goldman's clause is disjunctive (S's justification for believing *p* is undermined iff: S permissibly believes that her belief in *p* is not permitted *or* S impermissibly believes that her belief in *p* is non permitted *or* etc...). Extra disjuncts could easily be added to deal with new counterexamples⁶⁷, but it is not clear that this is a fully satisfying treatment of the intuitive, normative notion of justification.⁶⁸ However, further discussion of whether Goldman's account of justification captures the intuitive, normative, notion is unnecessary: for his definition is *not intended* to capture the intuitive notion. As we have noted, Goldman defines "justification" in a technical sense, clearly not intended to be even extensionally equivalent to the intuitive one (for in the technical sense justification is, while in the intuitive sense it is not, necessary for knowledge). The difference is clear: whether a subject's beliefs are justified in the intuitive sense has *methodological* significance for her; a subject's judgements about which of her beliefs are justified guide her in adjusting those beliefs. But Goldman's notion of justification plays no such role. It is simply a standard against which a third person can judge whether S's beliefs are the products of a reliable system of J-Rules. Mark Kaplan has argued, with some plausibility, that this leaves it very unclear why anyone should be interested in whether their beliefs are justified in Goldman's sense.

The problem is that, if it is characteristic of the nonregulative use of "justified" that it has no methodological import - that there is nothing in the set of rules to be consciously adopted by an inquirer that calls for her to determine whether any of her beliefs is justified in the nonregulative use of that term - then it is hard to see what point there could be to a system of evaluation dedicated to saying when it is that a person's belief is justified in the nonregulative sense. After all, in the sense of "justified" in question, it would seem that we might engage in inquiry as

⁶⁷ Goldman in effect does just this by introducing a distinction between *ex post* and *ex ante* justification to deal with a series of counterexamples from BonJour. I will pass over discussion of this. Goldman 1986: 111-3; 1992: 127-8.

⁶⁸ BonJour is the most trenchant critic. See his entry under "Externalism/Internalism" in Dancy and Sosa 1993 for a summary of objections to Goldman's account of justification; BonJour 1985 contains more in depth commentary.

Goldman's reliabilism

scrupulously and carefully as anyone can, yet never have occasion (and never suffer for our failure) even to *inquire* into whether our beliefs are justified!

Moreover [...] if it is characteristic of the nonregulative use of "justified" that it has no methodological import, it is hard to see on what basis one could judge [...] that the expression "justified," is being incorrectly applied.⁶⁹

While it is crucial for Goldman's naturalistic project that he give a non-normative analysis of the concept of justification, this project seems not to be served by his development of a technical notion of justification. The problem Goldman should be addressing, it seems, is that of how an apparently normative intuitive concept of justification can have a contingent relationship with knowledge, being neither necessary (nor, with truth and belief) sufficient for knowledge, but able in some cases through its absence to undermine it. We will return to this problem, and in particular the connection between justification (in the intuitive sense) and relevance, in the next chapter.

Our intention in examining Goldman's notion of justification was to see whether his justification clauses provided a mechanism which would prevent the occasional false belief produced by a locally reliable and appropriately produced and individuated belief forming process from being warranted. Our conclusion must be negative. Since Goldman defines "justification" by appealing to reliability, some false beliefs can clearly be justified (to this extent, at least, his notion of justification matches the intuitive one). This would not be a problem if some mechanism could be given such that the set of possible justified false beliefs was disjoint from the set of false beliefs produced by locally reliable, etc., processes, for then the conditions in Goldman's definition of warrant would interact to block the possibility of warranted false beliefs. But no such mechanism is offered.

Goldman's 1986 analysis of knowledge

We are now in a position to attempt to state the definition of knowledge we are attributing to Goldman. As he nowhere gives an explicit definition, we are forced

⁶⁹ Kaplan 1994: 353.

Goldman's reliabilism

to extrapolate one. We will do this by combining the various aspects of his account as follows:

S **knows** p iff:

p is true and S believes p and

- (i) S's belief in p results from an appropriate belief forming process α ;
- (ii) S's belief in p is justified.

A belief forming process α is **appropriate** iff:

- (a) α is globally and locally reliable;
- (b) α is the narrowest process causally operative in producing the belief in question;
- (c) α was acquired or is sustained by a second order belief forming process which produces reliable processes, or modifies existing processes so as to improve their reliability.

Belief forming process α is **locally reliable** iff

in most *relevant* situations: if α produces a belief in p then p is true.

Belief forming process α is **globally reliable**

iff in most situations: if α produces a belief in p then p is true.

S's belief in p is **justified** iff:

- (a) S's believing p is permitted by a right system of J-rules, and
- (b) this permission is not undermined by S's cognitive state at t .

A J-rule system R is **right** iff:

R permits certain psychological process, and the instantiation of these processes in normal worlds results in a truth ratio of beliefs that meets some specified high threshold.

Permission to believe is **undermined** iff:

S permissibly believes that her belief in p is not permitted *or* S impermissibly believes that her belief in p is non permitted *or* S believes that conditions C are not satisfied, where in fact the satisfaction of C is a necessary condition for S's belief in p to be justified *or* etc... .

Our requirements of a theory of knowledge were that it secures the infallibility of warrant; yields a definition of knowledge which excludes counterexamples based on external defeaters and internal defeaters; solves the generality problem; and explains the difference between deviant and nondeviant

Goldman's reliabilism

belief forming processes. Goldman's theory, as presented above, offers responses to many of these. Counterexamples based on internal defeaters are intended to be excluded by the no-undermining clause, while the generality problem and the problem of deviance in general are addressed by the specification of "appropriateness".

But as regards the requirements that warrant be infallible and that external defeaters be excluded, the definition is inadequate as it stands. Clauses (i) and (ii) say what it is for a belief to be warranted; the account of warrant will be infallibilist just in case it secures the truth of the warranted belief independently of the separate truth requirement with which the definiens begins. But this account of warrant does not secure the truth of the warranted belief, for appropriate production and justification (in Goldman's technical sense) are both fallible, and as we recently noted, there is no requirement that allows the two clauses to interact in securing truth.

The problem, of course, is the relaxed requirement for local reliability. By not requiring appropriate belief forming processes to be infallible in discriminating truth from relevant falsehood, Goldman loses the infallibility which we argued is built into the 1976 account. But this failure to define local reliability in terms of local infallibility just seems to be a slip on Goldman's part. In the next chapter, therefore, we will look at the prospects for strengthening the requirement of local reliability to bring it into line with the 1976 account, and of giving the definition of "relevance" with which that account needs to be completed. It will be suggested that the account of knowledge which arises is only dubiously naturalistic, and therefore only dubiously acceptable to Goldman.

CHAPTER EIGHT

Process reliabilism and contextualism

Authors in the tradition of explaining the a priori inaugurated by Kitcher have tried to develop naturalistic theories of a priori knowledge within a reliabilist framework drawn from Goldman. They have taken it to be obvious that a priori warrant, like warrant in general, is fallible, and hence that it is wrong to appeal to the concept of infallibility in explaining experience-independence. But in chapter six we saw an argument, based only on assumption of the closure principle for knowledge, that the property of warrant is infallible. If this argument is sound, it is indeed wrong to try to explain the notion of experience-independence by appeal to the notion of infallibility; however, the reason for this is the opposite of that assumed by Kitcher's naturalist critics, and moreover reliabilists must now face the problem of explaining just what infallible warrant is. In the previous chapter we argued that Goldman's 1986 theory does not require warrant to be infallible. If this is right, then Goldman's 1986 analysis of knowledge will face Gettier problems and does not provide a secure foundation for a naturalistic theory of a priori knowledge. However, we have also suggested that this problem could easily be fixed by reverting to the account of local reliability given in "Discrimination and Perceptual Knowledge", which apparently does require warrant to be infallible. The purpose of this chapter is to see how this modification can plausibly be implemented. Having done so, we will consider whether the resulting account meets the requirements of naturalism.

Our strategy for rendering Goldman's account infallibilist, then, depends on clause (i) of Goldman's definition. We redefine "local reliability" so as to bring it in line with the 1976 account, as follows:

Belief forming process α is **locally reliable** iff in all *relevant* situations: if α produces a belief in p then p is true.

On this account, "appropriate" belief forming processes never produce false beliefs. This cure might at first seem worse than the disease: our problem was to

Process reliabilism and contextualism

understand how warrant can be infallible, and the solution involves supposing that appropriate belief forming processes themselves never produce false beliefs. We noted in chapter six that there is an important difference between warrant being infallible and appropriate processes never producing false beliefs, and we suggested that to require the latter seems to lead to scepticism. The key to avoiding scepticism while requiring appropriate processes to be infallible is to tie "appropriateness" to the circumstances in which the belief forming process is employed. As we noted, this is precisely what Goldman's 1976 response to the fake barns problem seems to do: in case G, we suggested, the fact that S's visual processes can't distinguish real from relevant fake barns makes S's visual processes inappropriate in the circumstances. What we need to complete this account is a theory of relevance that explains this context dependence of appropriateness.

That our account of warrant would require attention to the circumstances of the knower was to be expected. Naturalists frequently stress the importance of the "situatedness" of cognitive subjects. Our knowledge-producing processes are designed by evolution to produce knowledge within environments with certain characteristics, and their ability to produce knowledge under *normal* conditions should not be taken to be undermined by the possibility that the processes go astray outwith normality.¹ For a warranting process to produce knowledge it must be operating within a favourable environment. The distinguishing features of favourable environments with respect to typically human knowledge-producing processes are myriad, including that there is adequate illumination; that the refractive index of the local medium for the propagation of light meets certain parameters; that local pressure and temperature is standard; that the environment supports generalisations based on sampling (that is, that the environment supports induction); that the local community of knowers is generally trustworthy, etc. Having recognised the importance of these conditions, it would be implausible to read the requirement that belief forming processes be reliable as a requirement that they produce a majority of true beliefs across all situations, since this would be to

¹ Cf.: Kornblith 1994, Craig 1990, Cohen 1988. This epistemological point has much in common with the increasingly popular view that our cognitive systems exploit environmental regularities as short cuts in processing (Clark and Chalmers 1998).

require that they produce true beliefs in abnormal situations. Hence, a distinction between relevant and non-relevant situations is also motivated.

However, in appealing to relevance a serious problem arises. The argument, given in chapter six, that warrant is infallible turned on the acceptance of the principle of the closure of knowledge under known entailment, schematically:

$$K\alpha, K(\alpha \models \beta) \vdash K\beta.$$

It would be extremely problematic if it turned out that the adoption of an account of knowledge which gives a central role to the notion of relevance is inconsistent with an adoption of the closure principle. If the closure principle is invalid, then we have in fact seen no cogent reason to think that warrant is infallible. Yet a number of authors have taken a "relevant alternatives" account of knowledge to indicate that the principle of closure is indeed invalid.² It seems very likely, however, that the closure principle must be sustained.³ I will first demonstrate the problem, and then propose the adoption of an account of relevance on which closure is sustained.

Closure and relevance

In his influential paper "Epistemic Operators" Dretske argues that the sentential operator "knows that ____" is "semi-penetrating".⁴ An operator ϕ is semi-penetrating iff, where ϕp , ϕ also applies to some but not all of the logical consequences of p .⁵ To avoid trivialising his thesis, Dretske adds the condition that the logical consequences in question must be known consequences; the claim he rejects, then,

² The *locus classicus* is Dretske 1970. Famously, Nozick (Nozick 1981) also denies closure, but as Nozick points out, his account draws heavily from Dretske. Goldman too says that his relevant alternatives account allows him to avoid sceptical arguments based on closure (Goldman 1986: 56).

³ See DeRose's "abominable conjunction" defence of the principle of closure: DeRose 1995: 27.

⁴ Dretske 1970.

⁵ Dretske contrasts semi-penetrating operators with "fully penetrating" operators such as "it is true that ____", "it is necessary that ____", and with "non-penetrating operators" such as "it is strange that ____", "it is accidental that ____", where the operator does not transfer even to the most immediate logical consequences of the proposition operated on.

Process reliabilism and contextualism

is that the epistemic operator "knows that___" penetrates to all the known logical consequences of a proposition. That is, by claiming that "knows that ___" is only semi-penetrating Dretske is denying the closure of knowledge over known entailment.

Initially, Dretske argues against the validity of the closure principle by exhibiting a scenario in which, he takes it, closure obviously fails. This scenario is as follows: S is in a zoo looking at some zebras. Let p be the proposition: the animal before S is a zebra; let q be the proposition: the animal before S is a mule cleverly painted to look like a zebra. Now, q is an alternative to p ; that is, p entails $\neg q$. Dretske maintains that, intuitively, S knows p ; however, to know $\neg q$, S would have to perform certain actions, gather certain evidence, and S has not done this: hence Dretske concludes that S does not know $\neg q$. Since S is in a position to know that q is an alternative to p , closure appears to fail. That is to say, the deductive closure argument

- (i) Kp
- (ii) $K(p \vdash \neg q)$
- so, (iii) $K\neg q$

seems to have been shown to be invalid. However, as he observes, this example only convinces if we are antecedently inclined to reject scepticism. A sceptic, on the other hand, could take the falsity of (iii) to transfer, via *modus tollens*, over to falsity of (i): in effect, arguing that S does not know that the animal before her is a zebra, on the basis of the fact that she does not know that it is not a cleverly painted mule. So, if we seek to convince the sceptic, we will need another, deeper, reason to think that closure fails.

Dretske tries to find a deeper reason by drawing analogies between the epistemic operator "knows that___" and other operators for which, he claims, the corresponding closure principles are not valid. The operators he discusses are " R is the explanation for ___", " R is the reason for S to ___", and the subjunctive conditional " R would not have been the case unless ___".⁶ But as Dretske himself points out, this

⁶ Dretske 1970: 1017.

line of argument could be thought weak, since it could be doubted that the logic of these operators is in the relevant respects the same as that of the epistemic operator. Moreover, it is not clear that Dretske manages to establish significant negative conclusions about the corresponding closure principles for these operators. As Dretske observes, given that we want to analyse knowledge counterfactually, the operator with the best *prima facie* chance of being relevantly analogous to the epistemic operator is the subjunctive conditional "*R would not have been the case unless* ____". Writing " $p \Rightarrow q$ " for "*p would not have been the case unless q*", the corresponding closure principle for this conditional is: $p \Rightarrow q, (p \rightarrow r): p \Rightarrow r$. Dretske proceeds by trying to give a counterexample to this inference pattern. It is, he thinks, easy to imagine a situation in which:

- (i) **The wall looks green (to S) \Rightarrow the wall is green**
 - (ii) **The wall is green *entails* the wall is not white cleverly illuminated to look green (to S)**
- are both true; yet, it is *not true* that
- (iii) **The wall looks green (to S) \Rightarrow the wall is not white cleverly illuminated to look green (to S).**⁷

But it is extremely difficult to determine whether Dretske can use this example to show that the closure principle for " \Rightarrow " is invalid, for it is extremely difficult to determine what are the semantics for the conditional "*p would not have been the case unless q*". The semantics for this conditional cannot be equivalent to the semantics for the normal counterfactual conditional "*if p had been true, q would have been true*", for if we interpret " \Rightarrow " as the normal counterfactual then the corresponding closure principle is *valid*.⁸ Dretske's counterfactual looks like a contraposed version of the normal counterfactual, but it cannot be, since contraposition is not a valid operation for counterfactual conditionals.⁹ So it is just not clear how to interpret Dretske's proposed counterexample to closure.

⁷ *Ibid.*: 1020. I have substituted the symbol " \Rightarrow " for Dretske's own " \rightarrow ".

⁸ Lewis 1996: 563 n.21.

⁹ For example: the counterfactual *if it had rained it would not have rained heavily* can be true; but its "contraposition", *if it had rained heavily it would not have rained* is absurd.

Process reliabilism and contextualism

In fact, the operator "*R is the explanation for* ___" seems to provide more secure grounds for the analogy Dretske wants to draw.¹⁰ Explanations seem typically to be *contrastive*: that is, we do not in general explain "why *p*" but: "why *p* rather than *q*". Dretske seems clearly right to say that we may have an explanation why *p*, and know that *p* entails *q*, but not have an explanation why *q*. For example, that *Brenda was full* explains why *Brenda did not order dessert*; and we know that *Brenda did not order dessert* entails that *Brenda did not: order dessert and throw it at the waiter*.¹¹ But that *Brenda was full* is not the explanation why *Brenda did not: order dessert and throw it at the waiter*. Perhaps we simply have no explanation why Brenda, in her rage about the service, did not take that step. But our inability to give an explanation here does not undermine the explanation of why Brenda did not order dessert. This introduction of the new possibility which we can't explain raises a new set of facts to salience, which our original explanation (that Brenda was full) cannot explain. But relative to the original set of salient facts, our explanation seems perfectly in order.

Dretske thinks that the logic of "knows that___" has the same structure as that of "*R is the explanation for*___". To say that *S* knows *p* makes tacit reference to a set of relevant alternatives which give content to the attribution of knowledge to *S*.

To know that *x* is *A* is to know that *x* is *A* within a framework of relevant alternatives, *B*, *C*, *D*. This set of contrasts, together with the fact that *x* is *A*, serve to define what it is that is known when one knows that *x* is *A*.¹²

A "relevant alternative", he writes, is an "alternative that might have been realised in the existing circumstances if the actual state of affairs had not materialized."¹³ Further, knowledge "penetrates only to those alternatives which form part of the

¹⁰ "*R is the reason for S to* ___" will also work in this regard.

¹¹ Dretske 1970: 1021.

¹² *Ibid.*: 1022.

¹³ *Ibid.*: 1021.

Process reliabilism and contextualism

network structuring the original context in which a knowledge claim was advanced."¹⁴

Dretske claims not to have given a precise definition of the notion of relevance.¹⁵ However, the most recent quotation suggests that Dretske must tie the notion of relevance very closely to the notion of penetration. Dretske says that knowledge penetrates only to those alternatives which are relevant, and it certainly seems that knowledge will penetrate to *all* those alternatives which are relevant. It follows that an alternative is relevant *just in case* knowledge penetrates to it. So it seems that Dretske is committed to the following equivalence (for propositions p and q):

$$p \text{ is relevant to } q \text{ iff: } Kp, K(p \vdash -q) \vdash K-q. ^{16}$$

If this definition is correct, then any theory of knowledge which appeals to a distinction between relevant and non-relevant alternatives will be automatically committed to rejecting closure. Where p entails $-q$, closure will fail in just those cases where $-q$ is not relevant to p .

However, as Stine points out, the appearance that the closure principle for knowledge fails is only sustained because Dretske is willing to evaluate Kp and $K(p \vdash -q)$ against a background of relevant alternatives which is different from that which licences the assertion of $-K-q$. But, as we have just seen, Dretske maintains that reference to the background of relevant alternatives is *part* of a knowledge claim, and hence that knowledge claims, even concerning the same proposition, are non-equivalent if they are evaluated against different backgrounds. Given this view of knowledge, it is illegitimate to claim that closure fails on the basis of cases which vary the background against which knowledge is attributed. As Stine writes, we cannot evaluate a deductive closure argument for validity without holding the set of relevant alternatives fixed:

¹⁴ *Ibid.*: 1023.

¹⁵ *Ibid.*: 1021 n.6.

¹⁶ This point is substantially similar to one made by Cohen 1988.

Process reliabilism and contextualism

This is as it should be; to do otherwise would be to commit some logical sin akin to equivocation. If the relevant alternatives, which have after all to do with the truth or falsity of the premises and conclusion [of instances of closure arguments], cannot be held fixed, it is hard to see on what basis one can decide whether the argument form is valid or not. And if the set of relevant alternatives is one thing for the first premise and another for the conclusion, how do we determine what it is for the second premise, and how does this affect the truth of the second premise?¹⁷

Given Dretske's definition of "relevance", an argument that the closure principle has not been shown to fail will also be an argument that the relation of relevance is simply equivalent to the relation of known entailment. Since it clearly is not so equivalent, we need to redefine "relevance". Stine offers the following condition for relevance:

(in situation *w*) an alternative *p* is relevant only if (in *w*) there is some reason to think that *p* is true.¹⁸

As we have noted, if the principle of closure is maintained, Dretske's zoo example becomes an instantiation of a general form of sceptical argument. The sceptic performs a *modus tollens* from the proposition that we don't know that sceptical scenarios do not obtain to the conclusion that we do not have knowledge of ordinary propositions.¹⁹ Stine wants to maintain closure, but does not propose to endorse scepticism. Instead she enjoins that, in the zebra case, the sceptic tries to *shift* the background of relevant alternatives by claiming that S does not know that the animal at which S is looking is a cleverly painted mule; however, confronted by a sceptic who claims this, S can rightfully refuse to allow the background of relevant alternatives to be changed. If S admits that she does not know that the animal is not a cleverly painted mule, she will have allowed the background of

¹⁷ Stine 1976: 256.

¹⁸ *Ibid.*: 252.

¹⁹ The alternative, to perform a *modus ponens* from the assumption that we know in ordinary cases, concluding that we know sceptical scenarios do not obtain, is often described as a *Moorean* response, after G.E. Moore's quick refutation of external-world scepticism.

relevant alternatives to shift, and against the changed background no longer knows that the animal before her is a zebra; if she refuses to let the background change, she continues to know that it is a zebra, and relative to that background she knows that the animal is not a cleverly painted mule.²⁰

Note that in advancing this solution, Stine has to reject Dretske's assumption that under normal circumstances *S* does not know that the animal before her is not a cleverly painted mule. But Dretske's rationale for this was plausible: it was that *S* has not, in those normal circumstances, performed any of the operations that would be required to know such a thing - she has not looked for paint pots, applied paint remover, etc. On Stine's account, where *q* is an alternative which is not relevant, $\neg q$ is known, but is known vacuously.²¹ This suggestion is certainly counterintuitive: some have found it totally implausible.²² However, it may be that the strangeness of claiming that we know irrelevant alternatives can be mitigated by other factors of our completed account. The account of relevance I will shortly endorse (Lewis's) retains Stine's proposal regarding irrelevant alternatives, but renders it more plausible.

Stine's treatment of relevance is, however, unsatisfactory. It is plausible that relevance is neither a purely psychological, a purely epistemic nor a purely metaphysical relation: alternatives can be relevant that we don't think are relevant (such as in the fake barns case), have no evidence are relevance, and perhaps even metaphysically impossible states of affairs can be relevant, if the believer does not realise that they are so impossible.²³ Stine's account seems to place disproportional weight on the epistemic aspect of relevance; accordingly, although we will endorse Stine's general strategy for combining a relevant-alternatives account of knowledge

²⁰ It is not, in fact, clear that Stine can consistently maintain this, for she suggests, tentatively, that the background of relevant alternatives is shifted by the very utterance of "*q*" in the second premise. If the mere utterance of a proposition is enough to change the background of relevant alternatives, any efforts on the part of *S* to "resist" the shift in background will be futile.

²¹ In Stine's formulation, non-relevant alternatives are known without evidence. I avoid Stine's own formulation since it sits uncomfortably with our previous endorsement of Williamson's evidence as knowledge thesis (Williamson 1997).

²² Cohen 1988.

²³ Suppose it is necessary that water is H_2O . We can imagine *S*, perhaps faced with a chemistry test, worrying whether water is H_2O , and thereby creating as relevant a situation in which water is not H_2O .

with acceptance of the closure principle, we still need to locate a satisfactory account of relevance.

Lewis's account of "proper ignoring"

The most detailed treatment of relevance currently available is given by Lewis in his 1996 paper "Elusive Knowledge". We cannot simply adopt Lewis's account *tout court*: one reason for this is that the theory of a priori knowledge which we will endorse affords a central place to the role of belief forming processes, and Lewis's account is not couched in terms of belief forming processes. The intention is to expropriate Lewis's account of relevance and combine it with the naturalistic account of knowledge derived from Goldman in order to create a working reliabilist account.

The basic idea behind Lewis's theory of knowledge is that a subject knows *p* just in case her evidence rules out the possibility that *p* is false. That is:

Subject *S* knows proposition *P* iff *P* holds in every possibility left uneliminated by *S*'s evidence; equivalently, iff *S*'s evidence eliminates every possibility in which not-*P*.²⁴

In practice, though, we just cannot eliminate all the possibilities: we have neither the time nor the cognitive resources to do a thorough job. This being the case, Lewis suggests, we allow ourselves to ignore many of the possibilities that really exist, the better to keep track of the ones we really have eliminated. To ascribe knowledge is in effect to say:

The possibilities eliminated, whatever else they may also include, at least include all the not-*P* possibilities; or anyway, all of those except for some we are presumably prepared to ignore just at the moment.²⁵

²⁴ Lewis 1996: 551.

²⁵ *Ibid.*: 563.

Process reliabilism and contextualism

But what a subject may be ignoring and still correctly be said to know depends on the context in which knowledge was attributed to her. This context-dependence is what will resolve the tension between the intuition that ordinary belief forming processes can produce false beliefs, and the suggestion that Goldman's notion of warrant can be made infallible by requiring that appropriate belief forming processes never produce a false belief. Lewis expresses this problem slightly differently, but the idea is the same. As he writes, intuitively

...it seems that knowledge must be by definition infallible. If you claim that S knows that *P* and yet you grant that S cannot eliminate a certain possibility in which not-*P*, it certainly seems as if you have granted that S does not after all know that *P*. To speak of fallible knowledge, of knowledge despite uneliminated possibilities of error, just *sounds* contradictory.²⁶

However, he goes on, minimal reflection reveals that our ordinary knowledge claims are indeed made against a background of uneliminated possibilities of error. Thus, either we must give up the idea that knowledge is infallible, or we must give up the idea that ordinarily we have knowledge. Neither option is appealing.

The solution to this paradox, roughly speaking, is that we retain the idea that possession of knowledge is incompatible with unexcluded possibilities of error, but allow that certain possibilities of error can be excluded not by ruling them out with evidence, but by *ignoring* them. The chief part of Lewis's theory of knowledge - and the part to be expropriated for our own purposes - concerns what may, and what may not, be properly ignored in a situation. To this end, Lewis proposes a series of rules. These are:

Rule of Actuality. The possibility that actually obtains may not be ignored. Hence, no false proposition may be presupposed²⁷.

Rule of Belief. A possibility that a subject believes to obtain, or ought to believe to obtain, may not be ignored.

²⁶ Lewis 1996: 549.

²⁷ We presuppose a proposition if we ignore all the possibilities in which it is false.

Process reliabilism and contextualism

Rule of Resemblance. If one possibility saliently resembles another, then: if one of them may not be properly ignored in virtue of rules other than the Rule of Resemblance, then the other may not.²⁸

Rule of Reliability. Defeasibly, we may ignore the possibility that our ordinarily reliable belief forming processes mislead us.

Rules of Method. Defeasibly, (i) we may presuppose that a sample is representative and (ii) we may presuppose that the best explanation of our evidence is the true explanation.

Rule of Conservatism. Defeasibly, we may ignore the possibilities that people around us commonly ignore; equivalently, we are permitted to adopt the usual and expected presuppositions of our community.

Rule of Attention. A possibility that is being attended to is not being ignored. Hence, if we are *wondering* whether we may ignore a possibility, we are attending to it, and so we are not ignoring it.

On a *contextualist* theory of knowledge, the *reference* of "knows" in an assertion of "S knows *p*" varies with the context of assertion. Two of Lewis's rules in particular render his account contextualist. First, the Rule of Attention requires that what a subject must be able to rule out in order for her to know depends on what situations she is directing her attention to: the more situations to which she is attending, the more she will have to be able to rule out in order correctly to be said to know. Second, by the Rule of Resemblance, the standards of similarity which must be met for a situation to become salient vary with the circumstances of attribution of knowledge: so, the standards for salience in force in ordinary circumstances may differ greatly from the standards in force in, say, a court of law.²⁹

I propose to read Lewis's rules for proper ignoring as rules governing what situations are relevant in a given context; I propose to adopt these as rules for relevance, and use them to complete Goldman's subjunctive account of reliability. So: in any given situation *a*, a possibility *w* is relevant just in case, in *a*, *w* is not properly ignored. Lewis's account sustains closure for much the same reasons as does Stine's, while giving a more filled-out and plausible account of the notion of relevance. For Lewis, the reference of "knows" depends on the context of

²⁸ This requirement is needed to prevent sorites reasoning to the conclusion that everything saliently resembles everything else, and hence preventing us from ignoring anything.

²⁹ Something like the Rule of Resemblance also seems to be central to Fogelin's contextualist account of knowledge (Fogelin 1994).

attribution of knowledge; deductive closure arguments can only be evaluated while the reference of "knows" remains fixed, as Stine pointed out, and this requires that we keep the background of relevant alternatives unchanged between evaluation of the premises and the conclusion. But, by the Rule of Attention, the mere mentioning of a situation makes it relevant. So, in Dretske's zoo case, the mentioning of the possibility that the animal before S is a cleverly painted mule raises it as a relevant unexcluded possibility of error, and while this possibility is not being properly ignored, S does not know that the animal before her is not a cleverly painted mule, and hence does not know that it is a zebra. But while the possibility is properly ignored, S knows she sees a zebra and so *also* knows, without having to check, that she does not see a cleverly painted mule. As we have noted, this is a slightly strange consequence; we will shortly look at the way Lewis manages to offset this strangeness.

Contextual reliability

We now look at how the theory of relevance drawn from Lewis could be combined with the reliable-process theory of warrant drawn from Goldman. The solution we will adopt is, roughly, that where we ascribe knowledge to a subject on the basis of a belief forming process α producing a belief in p , if we are permitted to ignore all the situations in which α produces false belief, then α will be warrant conferring. That is: when knowledge of p is attributed to S, the context of attribution determines a set of relevant alternatives, and if the attribution of knowledge succeeds, then in none of those situations does the belief forming process which produced S's belief in p produce a false belief. Warrant is infallible because appropriate belief forming processes don't produce false beliefs: scepticism is avoided by adding the requirement that whether a belief forming process is appropriate depends on the circumstances of ascription of knowledge.

We now have rules for determining what situations will be relevant in a context. As we have seen, these rules are context-sensitive in two ways: what is relevant depends on what the ascriber is (and ought to be) attending to, and on

what standards of similarity are in force. Now, Goldman's theory is messy: so messy, in fact, that we might be justified in rejecting it even if we are right to think that the analysis gives an extensionally correct account of knowledge; for we might think that a successful analysis of a concept would not just be extensionally equivalent to the original concept, but would meet certain standards for intensional equivalence also.³⁰ Fortunately, the adoption of Lewis's rules as rules for relevance allows us to simplify Goldman's account considerably. For the rules for relevance double up many of the clauses of Goldman's definition of knowledge.

The rule of actuality requires that truth is always relevant: so we can drop the extra truth clause in Goldman's definition.³¹ Further, if we adopt the Lewisian theory of relevance we do not have to include into the definition of warrant a separate clause requiring that an appropriate belief forming process is produced by a reliable second-order process (a process which produces belief forming processes - see chapter seven). This is because where knowledge is attributed to a subject whose belief was produced by an unreliably produced belief forming process, the fact that the process was unreliably produced will create relevant uneliminated possibilities of error.³²

For example, recall Goldman's case of Sam and the twins, Judy and Trudy.³³ In this scenario, Sam has a set of perceptual cues for recognising Judy and Trudy, and up to a certain time t these sets have been indistinct, leaving Sam unable to distinguish Judy from Trudy. At t Sam hits his head and a new feature is added to his set of cues for recognising Judy. From t onwards, he will believe someone to be Judy only if they meet all the previous conditions Sam used to judge to be true of Judy, plus they have a mole over their right eyebrow. But at t , Judy happens to develop such a mole. So Sam can now distinguish her from Trudy, but his ability to so has been brought about so serendipitously that his first recognition of her will not count as knowledge. Given our knowledge of the way Sam's ability to distinguish Judy from Trudy was brought about, we should consider a cluster of

³⁰ Cf. Craig 1990, Williamson 1995.

³¹ Cf. Lewis 1996: 554.

³² Goldman mentions this sort of solution for dealing with deviant production, and passes over it without comment: Goldman 1976: 789.

³³ See chapter seven.

possibilities relevant: in one, Judy doesn't develop a mole; in another, Trudy also develops a mole; in a third Sam is rendered less able to distinguish Judy from Trudy by the bump on the head (perhaps it leaves him extremely short-sighted). Sam can exclude none of these, and none is properly ignored; therefore it would be wrong to attribute knowledge to Sam.

It seems plausible that related reasoning would hold for the other case of unreliable second-order production we have seen, the case in which S is taught a reliable algorithm by an unreliable quack mathematician.³⁴ So Goldman's 1986 definition can be further simplified: we do not need a separate clause requiring reliable second-order production.

A third area in which we could seek simplification is with respect to Goldman's justification clauses. For Goldman, justification is necessary for knowledge. His main clauses for justification are: S's belief in p is justified iff: (a) S's believing p is permitted by a right system of J-rules, and (b) this permission is not undermined by S's cognitive state at t . It seems likely that these clauses can be assimilated to the Rules for relevance. Clause (a) is, in effect, that S's belief is produced by a process which is globally reliable (that is, it produces a suitably high proportion of true beliefs overall); where S's belief is produced by a process that is not globally reliable, it is clear that one of the Rules will operate to block the attribution of knowledge to S.³⁵ Clause (b) is, as we saw, disjunctive, and it is not clear whether Goldman intends it to be complete as it stands, or whether he would allow further disjuncts to be added to deal with new counterexamples should they arise. One way in which S's permission to believe $-p$ can be undermined is if S actually believes $-p$; here the rule of Attention operates to block correct attribution of knowledge to S. Similar mechanisms can probably be found which would assimilate Goldman's other undermining clauses to the Rules for relevance.

However, as we have noted, it is not at all clear why Goldman introduces a requirement of justification in this technical sense. Justification is naturally taken to be normative, and in this intuitive sense is neither necessary nor sufficient for

³⁴ Case J from chapter seven.

³⁵ For illustration, see the contextualist treatment of the astrology case, case E: page 189, this chapter.

Process reliabilism and contextualism

knowledge.³⁶ Lewis's Rules do include clauses which deal with justification in the intuitive sense: the Rule of Belief requires that a possibility that S ought to believe obtains must be considered relevant. But this means that Lewis's account of relevance is normative: it therefore cannot be acceptable to Goldman as it stands.

We will discuss this problem in more depth shortly; perhaps, though, Goldman could drop the normative aspect of the Rule of Belief and hope that what remains of his definition of knowledge contains the machinery needed to deal with the cases that the absence of the normative concept leaves open. Assuming that he can, it seems that all we need in the definition of knowledge is the requirement that the process in question is the narrowest process causally relevant to producing S's belief, and that the process produces no false belief in any relevant alternative situation. This suggests a comprehensive rewriting of Goldman's account, which keeps the spirit of the reliable-process theory, but packs many of his extra conditions into the relevance-clause. The process-reliabilist definition of knowledge I suggest Goldman is left with is, then, as follows:

- (Ω) S **knows** *p* iff: S believes *p* and S's belief in *p* results from a belief forming process α such that:
- (i) α is the narrowest process causally operative in producing the belief in question;
 - (ii) α produces no false belief in any situation made relevant by the context of attribution of knowledge to S (where the relevant situations are determined by Lewis's Rules).³⁷

The contextualised account of reliability explains why S does not know in the Gettier-style cases we have seen so far, and also deals with some cases that we have not yet seen. From chapters six and seven, cases D-G are dealt with as follows.

- D S is in a room with two other people. She has known one of them for years, and knows that he has always loved Ford cars, that he has always bought them over other cars, that he has never expressed any interest in driving a

³⁶ Lewis also makes this point: Lewis 1996: 556.

³⁷ In other words, α is locally reliable (1976 version). On this account, the following property of a belief is infallible: "___ was produced by an appropriate belief forming process". The key is that whether a belief forming process is appropriate depends on the context.

Process reliabilism and contextualism

non-Ford car, etc. S also knows that yesterday this friend arrived at work in his own Ford car, and that he drove into work today (though she didn't see him arrive). S has, therefore, a justified belief that her friend owns a Ford car: but in fact he does not, having sold it just that morning. S does not own a Ford car. However, the second person, who is a stranger to S, does own a Ford car. Thus, S has a justified and true belief that *someone in the room owns a Ford car*.

Solution: the situation w , in which S's friend has just sold his Ford, while the stranger does not own a Ford, is a relevant possibility. This is because actuality is always relevant, and w saliently resembles actuality: in respect of S's friend it resembles actuality exactly; and in respect of what S knows about the stranger it is also exactly alike. Hence w may not be ignored, and in w S's inferential process produces a false belief: so clause $\Omega(ii)$ fails, and it is not correct to attribute knowledge of p to S.³⁸

E S uses astrology to form some belief p . p is in fact true, and S has complete subjective justification for her belief, having been convinced of the efficacy of astrology through being raised in an community where astrology was common, having seen apparently reliable results many times before, being versed in the underlying pseudo-theory, etc.

Solution: the possibility, w , that astrology leads S to believe $\neg p$ is relevant. Since astrology does not hook onto any epistemically significant regularities in nature, we can imagine the astrological divination process producing the negation of the belief it actually produces without altering any significant aspect of the rest of the world; hence there is a possibility, w , which resembles actuality in all respects bar S's belief; w saliently resembles actuality and actuality may never be ignored. Hence w may not be ignored, and in w the divination process produces a false belief. So $\Omega(ii)$ fails, and it would be incorrect to attribute knowledge of p to S.

F S has a certain belief p about her early childhood, but her parents tell her that her belief in p is false, and that her memory of the occurrence of p is

³⁸ Lewis 1996: 557.

Process reliabilism and contextualism

mistaken. Despite the attempted deception, S continues to believe in p . S's belief is true and her memory is working perfectly.

Solution: by the Rule of Belief, S ought to believe that the possibility, w , in which her parents are telling the truth obtains: hence w is relevant, and in it S's memory processes produce a false belief. So $\Omega(\text{ii})$ fails and it is not correct to attribute knowledge of p to S.³⁹

G S is driving along a road, and believes that she sees a barn, based on the fact that she seems to see the front of a barn that she is passing, and that she has no reason to believe that visual conditions are abnormal. Visual conditions are in fact normal, and in fact there is a real barn before her. However, unknown to S she is in a region of the country where many fake barn façades have been erected, such that if S had been looking at a barn façade she would still have believed she was looking at a barn, and it is pure luck that S is now looking at a barn.

Solution: the possibility, w , in which S sees a fake barn saliently resembles actuality in respect of the probability that S sees a fake barn, and so w cannot be ignored. In w , S's belief that she sees a barn is false. So, once more, $\Omega(\text{ii})$ fails: S's visual processes are not reliable in this context and it would not be correct to attribute knowledge to S.

The contextualist version of Goldman's reliabilism can also solve other problems which affect fallibilist versions of reliabilism. For example, the lottery paradox⁴⁰: S does not know that her lottery ticket will lose, even though, by increasing the number of tickets in the lottery, we can make her chances of winning arbitrarily low. Suppose S has a friend who is generally trustworthy; intuitively, S can gain knowledge through testimony from her friend. However, we may so reduce the chances of S winning the lottery that the probability of her winning is

³⁹ This solution is not available to Goldman, since we are assuming that he should drop the normative element of the Rule of Belief and hope that its work can be done by other aspects of the definition. It seems clear that in this case at least the hope will be born out: by the Rule of Attention, her parents' attempted deception raises to salience a possibility of error that S cannot exclude.

⁴⁰ Cohen 1988, Lewis 1996.

Process reliabilism and contextualism

lower than the probability that her friend is lying (we may suppose that the chance the friend is wrong is $1/10^4$, while the chance of winning the lottery is $1/10^7$). An account which ties warrant purely to metaphysical probability will have difficulty explaining how in the lottery case S does not have knowledge, while in the testimony case she does. But on the contextualist account matters are straightforward: for each ticket, there is a possibility in which it wins; before the winner is drawn, each of these possibilities resembles each other perfectly as far as S is concerned; so either all of these possibilities are relevant or none is; but one of the possibilities is actual, and may not be ignored; so none of them may be ignored; hence S does not know she will lose.⁴¹ In contrast, the rules of Reliability, Method and Conservatism allow S to ignore, albeit defeasibly, the possibility that her friend is mistaken, and so her fallible friend can be a source of knowledge.

Finally, we return to the sceptical argument which Dretske's zoo example instantiates, once the principle of closure is sustained. The general form of the argument is given by DeRose (for some ordinary proposition O (in Dretske's case: I see a zebra), and some sceptical proposition H (in Dretske's case: I see a mule cleverly painted to look like a zebra)):

- (i) I don't know not-H
 - (ii) If I don't know not-H, then I don't know O.
- So, (iii) I don't know O.⁴²

According to Lewis's rules for proper ignoring, the very mention of H in (i) brings H to our attention; hence by the Rule of Attention H is not properly ignored: hence H is relevant. And while H is relevant, (iii) is true: for there is a relevant situation in which whatever process generates our belief in O produces a false belief, hence

⁴¹ This solution from Lewis 1996: 557.

⁴² DeRose 1995: 1.

Process reliabilism and contextualism

$\Omega(ii)$ fails, and that process is not reliable in the context generated by our attribution of knowledge.⁴³

As we have noted, Lewis's solution mirrors Stine's in this respect: while H is not relevant, O is known and hence, by closure, H is known. But as noted in connection with Stine's account, it can seem counterintuitive to hold that possibilities can be known while they are ignored. Lewis's account of contextualism, however, addresses this problem. Lewis points out that knowledge gained through ignoring possibilities is absolutely tenuous: merely mentioning the possibility serves to dispel the knowledge.

Do I claim you can know *P* just by presupposing it?! Do I claim you can know that a possibility *W* does not obtain just by ignoring it? Is that not what my analysis implies, provided that the presupposing and the ignoring are proper? Well, yes. And yet I do not claim it. Or rather, I do not claim it for any specified *P* or *W*. I have to grant, in general, that knowledge just by presupposing and ignoring *is* knowledge; but it is an *especially* elusive sort of knowledge, and consequently is it an unclaimable sort of knowledge. You do not even have to practise epistemology to make it vanish. Simply *mentioning* any particular case of this knowledge, aloud or in silent thought, is a way to attend to the hitherto ignored possibility, and thereby render it no longer ignored, and thereby create a context in which it is no longer true to ascribe the knowledge in question to yourself or others. So, just as we should think, presuppositions alone are not a basis on which to *claim* knowledge.⁴⁴

In general, the more possibilities we can rule out, and the less possibilities we have to exclude by ignoring, the more robust our knowledge will be. Thus, relying on testimony from some authoritative source may be an efficient way of gaining knowledge, but the knowledge so gained can easily be undermined. By reflecting on and finding out more about what the authority said, we are able to rule out more possibilities for ourselves, and so our knowledge becomes more stable. This is a consequence of Lewis's account⁴⁵, and also seems to reflect the behaviour of

⁴³ The form of the example is taken from DeRose 1995, the solution from Lewis 1996. DeRose's own solution is analogous to Lewis's; but Lewis's account has the advantage that it is meant as a general theory of knowledge: DeRose's account is more narrowly directed against scepticism.

⁴⁴ Lewis 1996: 562.

⁴⁵ *Ibid.*

Process reliabilism and contextualism

knowledge in real situations. Knowledge through presupposing and ignoring seems to be a limiting case of this phenomenon: it is maximally unstable knowledge. The apparent continuity between the everyday phenomenon of the instability of knowledge, and the strange phenomenon of knowledge through ignoring, helps to render the latter less strange.

Problems for contextualism

The contextualist strategy has a lot to recommend it, and gives a plausible account of the knowledge that the closure principle requires we have of presupposed propositions. Further, by allowing simplification of Goldman's account it deflects much of the force of the criticism that his account does not capture the intuitive intension of the concept of knowledge. But contextualism is not without problems, and these should not go unremarked. I do not propose to give fully worked-out solutions to these problems; some key issues and problems, with sketches of suggested solutions where possible, are as follows. We will conclude with a discussion of whether the contextualist strategy is really open to naturalists.

(1) As we have already seen in the case of knowledge through presupposing and ignoring, the adoption of contextualism requires that we slightly adjust some of our peripheral intuitions about knowledge. Some further adjustments are also required. Lewis considers the question of how it is possible to act rationally without knowledge: imagine two epistemologists out on a walk while discussing distant uneliminated possibilities of error which, since by discussing them they are thereby attending to them, undermine all their knowledge.⁴⁶ Nevertheless, they continue to be able to navigate on the walk, and since they are able to navigate, it seems apt to say that they know where they are. Lewis suggests that we adopt the view that subjects are compartmentalised, allowing us to distinguish between what the philosophical compartments of the two walkers know, and what the navigation

⁴⁶ *Ibid.*: 564.

Process reliabilism and contextualism

compartments know. If we allow that it is only their philosophical compartments which are attending to distant uneliminated possibilities of error, we can allow that their navigation compartments still know where they are. This raises the question of what the whole compartmentalised thinker knows; a natural response would be to say that the overall thinker knows p just in case one of her compartments knows p , but as Lewis points out, this will raise new doubts about the principle of the closure of knowledge over known implication. For, one of S's compartments may know p , another know p entails q , and no compartment know q .⁴⁷ Lewis's preferred solution is to treat each of the overall thinker's compartments as a different subject, and so to say that the overall compartmentalised thinker both knows and does not know where they are. On this view, overall thinkers are treated as composed of many subjects; whether this idea is ultimately plausible will require further work.

(2) As Lewis points out, the Rule of Resemblance requires a seeming *ad hoc* exception.

We must apply the Rule of Resemblance with care. Actuality is a possibility uneliminated by the subject's evidence. Any other possibility W that is likewise uneliminated by the subject's evidence thereby resembles actuality in one salient respect: namely, in respect of the subject's evidence. That will be so even if W is in other respects very dissimilar to actuality - even if, for instance, it is a possibility in which the subject is radically deceived by a demon. Plainly we dare not apply the Rules of Actuality and Resemblance to conclude that any such W is a relevant alternative - that would be capitulation to scepticism. The Rule of Resemblance was never meant to apply to *this* resemblance! We seem to have an *ad hoc* exception to the Rule, though one which makes good sense in view of the function of attributions of knowledge. What would be better, though, would be to find a way to reformulate the Rule so as to get the needed exception without *ad hocery*. I do not know how to do this.⁴⁸

This seems unfortunate. One way of responding to it might be to appeal to Williamson's thesis that one's evidence consists of all and only the propositions that

⁴⁷ *Ibid.*

⁴⁸ *Ibid.*: 556-7.

one knows. The problem arises in the first place because Lewis allows that S's evidence can be the same in two situations, W_1 and W_2 , which are radically different in terms of what S knows. Similarity in point of evidence is meant to make the radically different situation relevant, and so create it as a possibility of error which cannot be excluded. The *ad hoc* exception to the Rule of Resemblance is just that we assert that resemblance between W_1 and W_2 in point of evidence alone need not make these two situations relevant to each other. But this *ad hoc* exception might not be needed if we accept the evidence=knowledge thesis. If evidence=knowledge, the fact that S's evidence is invariant across W_1 and W_2 entails that S's knowledge is invariant also. If S's knowledge is invariant then all the propositions that S knows in W_1 are true in W_2 ; therefore W_2 does not represent an unexcluded possibility of error.⁴⁹

This solution, however, is bought at the cost of the possibility of giving a reductive analysis of knowledge in Lewis's terms. For Lewis, S's evidence is identified with the S's total perceptions and memory at a given time: S's evidence is the same between two situations just in case S's total perceptions and memory are the same in those situations.⁵⁰ Lewis wants to use this concept of evidence as part of a reductive definition of the concept of knowledge, and so cannot adopt the evidence as knowledge thesis on pain of circularity. It may be, however, that Lewis's account does not avoid circularity in any case: we will shortly discuss whether the rules for relevance can properly be considered able to contribute to a reductive definition of knowledge. Putting this on one side, though, the evidence as knowledge thesis still seems unable to solve the problem for the Rule of Resemblance, since the two situations - actual, and sceptical scenario - will still resemble each other in point of evidence in Lewis's sense. Ultimately, then, it is not really clear how to respond to the problem without the *ad hoc* exception.

(3) Schiffer has objected that the semantics of knowledge attributions implied by contextualism is at odds with the way such attributions ordinarily work.⁵¹ We

⁴⁹ Williamson 1995.

⁵⁰ Lewis: 1996: 553.

⁵¹ Schiffer 1996.

Process reliabilism and contextualism

can illustrate the problem by returning to the sceptical argument form instantiated by Dretske's Zoo scenario.

- (i) I don't know -H
- (ii) If I know O then I know -H
- (iii) I don't know O.

As we've seen, according to contextualism, the sceptical problem posed by this argument is dissolved once we realise that the standard required to have knowledge (that is: the background of relevant alternatives against which our processes must be infallible if they are to warrant the beliefs they produce) shifts between (i) and (iii). (ii) is true for any fixed standard; and if we hold the standards fixed, either (i) will be false (for a relaxed standard) or (i) and (iii) will both be true (for a strict standard) - though once the standard relaxes again, we will regain our knowledge of O. Contextualism explains how the standard can shift: it is the mere mentioning of -H at (i), which, by the Rule of Attention, makes a situation in which H is true relevant, and so creates it as a possibility which cannot be excluded.

Schiffer holds that this account fails to explain why the sceptical paradox seemed paradoxical in the first place. Numerous expressions have context-dependent meaning, but fluent speakers have no difficulty in recognising the context-dependence, nor in grasping what context is in force on a given occasion of utterance. This being so, it is at least odd that - if the contextualist is right - the context dependence of knowledge has gone unnoticed for so long, and that speakers have so much difficulty realising what context is in force when utterances like (i) and (iii) are made.

What the contextualist needs to give is a fully worked out account of the semantics of knowledge attribution sentences which captures the indexical nature of such sentences, and explains how fluent speakers can be confused by them. Schiffer considers three ways in which such an account might be developed.⁵²

Hidden indexical account. This assimilates knowledge attribution sentences to sentences such as "it's raining" or "he's short" which contain "unarticulated

⁵² *Ibid.*: 326.

constituents"⁵³. These are "propositional constituents which aren't part of the semantic values of any terms in the uttered sentence."⁵⁴

Thus, an utterance of "It's raining" might express the proposition that it's raining in London, and an utterance of "He's short" might express the proposition that so-and-so is short for an NBA centre.

If we adopt the hidden indexical account, we must take it that sentences with the surface form "S knows *p*" express propositions of the form "S knows *p* relative to background of relevant alternatives N".⁵⁵

Indexical verb account. This builds the context-sensitivity of knowledge attribution sentences directly into the verb "to know". On this account, the verb itself is indexical; an assertion of "S knows *p*" expresses the proposition "S knows_N *p*", where N gives the background of relevant alternatives against which S is said to know and is determined by the context of utterance.

Vagueness account. Schiffer takes it that the verb "to know" is vague, and that the degree of vagueness of a term can vary with conversational purposes.

The verb "to know", like virtually every expression, is vague, and there is a certain context variability inherent in vagueness. The penumbras of vague terms can dilate or constrict according to conversational purposes.⁵⁶

None of these accounts, he holds, give a satisfactory explanation of how fluent speakers could be unaware of the context-dependence of knowledge-claims. Each account seems to entail that the context-sensitivity be obvious to the speaker; since it is not obvious, Schiffer argues, we must reject these as accounts of the semantics of knowledge attribution sentences. Schiffer infers that the contextualist account is unworkable, and concludes that there is *no* solution to sceptical paradoxes such as the one above. These paradoxes, he holds, arise because of a

⁵³ Perry 1986.

⁵⁴ Schiffer 1996: 326.

⁵⁵ This formulation is changed from Schiffer's: Schiffer writes that the proposition expressed is: "S knows *p* relative to standard N", for he is discussing a version of contextualism - DeRose's - which differs from the account offered by Lewis.

⁵⁶ *Ibid.*: 327.

deep-seated incoherence in our concept of knowledge. One element of the concept of knowledge allows us to take the evidence of our senses as constituting knowledge; but another element, acquired late in cognitive maturity, requires that we cannot object to the sceptic by appealing to the very knowledge that the sceptic calls into question.⁵⁷ Sceptical scenarios have force because they bring these two elements into acute conflict.

Schiffer's objection can perhaps be challenged. The key point he makes is that contextualism implausibly requires that fluent speakers be unable to recognise the context-sensitivity of knowledge attribution claims, nor be able to determine what context is in force on a given occasion of attribution. But we should note we do not require that fluent speakers be absolutely infallible with respect to their own language. Consider the following sentence: *the father of the Pharaoh's daughter is the mother of the Pharaoh's son*. Most people have difficulty in recognising that there is a reading of it on which it is trivially true.⁵⁸ There are many other related examples.

The analogy between lexical illusions and sceptical scenarios is not perfect by any means: a key difference concerns the phenomenology associated with having the solution to the problems pointed out. In the case of lexical illusions, once the correct reading is indicated, it is normal for all sense of mystery to vanish. This is not so clearly the case with sceptical scenarios: one can be led to accept a solution, but still feel a "pull" towards scepticism when the scenario is described.

One way in which the objection to our response to Schiffer might be met is to suggest that our capabilities themselves can be undermined in certain contexts: excessive attention to how we perform an action can leave us temporarily unable to perform it. Epistemology classes are contexts in which we focus our attention on the way we understand attributions of knowledge: perhaps this attention is

⁵⁷ Schiffer suggests that it is this bar against begging the question against the sceptic which explains why we are inclined to agree that we do not know that sceptical scenarios do not obtain. In claiming this he breaks with analyses which explain the lack of knowledge by reference to the *insensitivity* of one's belief that the sceptical scenario does not obtain (S's belief in *p* is insensitive iff S would believe *p* even if *p* were false: Cf.: Nozick 1981, DeRose 1995). Schiffer points out that the appeal to insensitivity seems insufficient to explain why we do not know that sceptical scenarios do not obtain, for let *p* be the following proposition: I am not a Brain in a Vat being caused to hallucinate that I am on a Spanish cruise. I know *p*; and yet my belief in *p* is insensitive: I would still believe *p* even if it were false.

⁵⁸ The father (of the Pharaoh's daughter) is the (mother of the Pharaoh's) son.

Process reliabilism and contextualism

excessive, and in these contexts our ability to spot the "correct reading" of sceptical scenarios is undermined. This might explain why the sceptical scenarios still feel compelling even after the contextualist solution has been pointed out to us. Outside the context of the epistemology class we no longer focus our attention on the way we use the concept of knowledge and, as Hume pointed out, are no longer troubled by scepticism.

These suggestions are very vague. A full response to Schiffer would require giving a detailed semantics for knowledge attribution sentences, and I am unable to do such a thing here. However, very tentatively, I suggest that the preceding observations regarding context-sensitivity of abilities and the existence of linguistic illusions leave us with some reason to think that such an account might be developed.

Nonreductive naturalism

A final doubt we must consider concerns whether the account of relevance developed here is naturalistically acceptable. We have already noted that the Rule of Belief requires that what a subject *ought* to consider relevant may not be properly ignored, and that this renders Lewis's account of relevance unacceptable to Goldman as it stands. However, it seems reasonable for the naturalist to hope that the normative clause in the Rule of Belief can be dropped and the strain taken by other parts of the definition of knowledge. But this worry is comparatively minor. A deeper worry concerns whether we have *really* excluded the concept of knowledge in giving our contextualist process-reliabilist analysis: although the concept of knowledge does not explicitly appear in the analysans of (Ω), there is still a nagging worry: perhaps we don't really understand the analysans any better than we understand the analysandum.

Consider, once again, the fake barns case: we can imagine a series of cases, ranging from one in which S is looking at a real barn and there is no nearby (counterfactually speaking) situation in which the barn she is looking at is fake; and

Process reliabilism and contextualism

ascending to case G, where S is looking at a real barn but there is a very close possibility in which the barn is fake. At the end of the series corresponding to case G, we deny that S knows; at the other extreme, we allow that S knows; at some point as we ascend the series we cross a threshold beyond which the possibility that she sees a fake is sufficiently high for us to deny, by appeal to the Rule of Resemblance, that she has knowledge. It does not seem possible for us to determine where that point is with any accuracy, but this reflects a vagueness in our attributions of knowledge and is not serious. However, it seems that the analysis does no work in reducing this vagueness. Presented with a case taken from some point on the series, we consult our intuitions about *knowledge* to decide whether S knows in that case, and it is based on our intuitions about knowledge that we decide whether the fake barn situation is relevant in that case. The intuitions about knowledge are primary.

The problem here is not that someone who was not versed in a relevant-alternatives account of knowledge would be able to pass judgement on whether S knew in some case drawn from the series without the benefit of being able to apply the concept of relevance. Such a consideration does not constitute a serious objection to a proposed analysis, for analyses of concepts are not in general meant to be easier to work with than the concept unanalysed. There are two standard motivations for conceptual analyses: one is a hope for increased precision concerning the conditions of application of the target concept; the other is a desire for reduction of the concept in terms of some privileged set of terms. The problem is that the analysis of knowledge in terms of relevance seems to do neither of these. The consideration of the series of barn cases suggests that the analysis does *no work* in clarifying the boundaries of application of the concept of knowledge. I can think of no cases where my indecision over whether the concept of knowledge applies can be resolved by appeal to the analysis in (Ω). Worse, it is simply not clear that the account is reductive. Many of the Rules for relevance are hedged, that is, they contain clauses specifying that they only obtain defeasibly. Further, the Rule of Resemblance appeals to *salient* resemblance, and standards for salience are allowed to vary. But we are not given any rules governing the application of defeasibility

Process reliabilism and contextualism

conditions, or the variation of standards for salient resemblance; and it might be suggested that no reductive clarification *can* be given on these issues: that is, that our ability to apply the Rules presupposes that we grasp the concept of knowledge. If this is the case then an analysis of knowledge in terms of relevance will not, in fact, be reductive. Although the concept of knowledge does not explicitly appear in the analyses, it is there implicitly: the Rules tacitly appeal to the concept of knowledge in the idea of defeasibility and shifting of standards for salience, and cannot be used to reduce that concept.

Faced with this, it seems that reliabilists must choose between four options, none of which is palatable. The first option would be to abandon the attempt to give a relevant alternatives account of reliability, and to find some altogether different solution to the problem of external defeaters. It is not at all clear what this solution would be.

The second option would be to maintain that the concept of relevance is in fact suitable for giving a reductive account of the concept of knowledge. This would require working to show that the normative element in the Rule of Belief can indeed be dropped, and that the application of the hedged Rules is governed by suitably non-epistemic, non-normative principles. It is not clear whether there is any serious prospect for success here; the chief reason for adopting this option would be dismay at the other prospects on offer.

The final two options both involve resigning oneself to the impossibility of providing a reductive definition of knowledge. We should note that to accept that the concept of knowledge cannot be reduced is not automatically to give up on the possibility of providing a naturalistic account of it. For it is a brute fact that not all concepts must be reducible if they are to be legitimately employed. To deny this would be to commit oneself to a vicious infinite regress of conceptual analyses. Botterill and Carruthers suggest that the implausibility of requiring reduction in general motivates the view that the concepts of natural science do not have to be reduced to be acceptable.

[W]e should accept that the existence of a variety of special sciences is a permanent, irreducible, part of our world view, reflecting the way in which the

Process reliabilism and contextualism

natural world is organised in terms of laws and principles operating at different levels of generality. And then all we need to do in order to *naturalise* some property, is show that it figures in the laws of some or other special science, in whose persistence we have good reason to believe.⁵⁹

Recall the two parameters governing what can be treated as scientifically acceptable, given in chapter four. To adopt Botterill and Carruthers's position is to take a permissive stance on the first parameter: it is to accept properties of high level sciences without requiring their reduction. However, a permissive stance here will not be sufficient to allow a nonreduced concept of knowledge to be naturalistically acceptable, for it seems plausible to say that knowledge is not a property which figures in the laws of any natural science. If we are to render the idea of unreduced knowledge naturalistically acceptable we will have to take a permissive stance on the second parameter, and allow that knowledge, though not currently part of any scientific theory, has the potential to be so. But, as we have already noted, longstanding metaphysical commitments bar us from allowing unreduced appeal to normative concepts in science, and an unreduced concept of knowledge would be norm-laden. The third option available to reliabilists, then, is to reconsider the injunction against allowing normative concepts in science.⁶⁰ If this seems too implausible, the only remaining option seems to be to eliminate the concept of knowledge entirely. But this final option is, in effect, an abandonment of the project of the naturalist project to reconcile denial of the autonomy of philosophy from science with acceptance of a meliorative role for epistemology.⁶¹ The eliminativist strategy seems to be nothing less than adoption of the idea of epistemology as pure description, an idea that even Quine, its progenitor, finds unappealing.

I cannot adjudicate between these four possibilities here; nor is it clear which way the traditional naturalist should turn. The traditional naturalist must certainly

⁵⁹ Botterill and Carruthers 1999: 186.

⁶⁰ Biology is the most plausible candidate for a science which deals with normative concepts. Biological norms are standardly analysed reductively using the concept of natural selection: a trait has a *function* just in case there has been selection for that trait to perform that function in the past. Whether such a reductive account is tenable is controversial (Cf. Walsh 1998, 2000).

⁶¹ See chapter four.

reject the fourth option, since they want to retain the meliorative role of epistemology; the second option is most in accord with their principles, but we have not seen that it has any obvious chance of success; to adopt the first option is really to resign oneself to despair; and it is not clear where the third option will lead. For the purposes of this thesis, however, we should note that only the fourth option gives us reason to doubt the compatibility of a priori knowledge and naturalism. The fourth option requires the rejection of a priori knowledge because it requires the rejection of knowledge itself; but while the traditional naturalist holds out the hope of giving a naturalistic analysis of knowledge, they seem to have no special reason to reject a priori knowledge.

Conclusion

Once we had established that all warrant, a priori or not, is infallible, the challenge became to show that a naturalistic account of knowledge could be given at all. This chapter, and the preceding one, have discussed the problems associated with this project. We have been seeking to take the core features of Goldman's reliable process theory of knowledge, augment them with an account of relevance, and so arrive at an acceptable theory. It turned out that the determination of what is relevant depends on the context of attribution of knowledge, and we raised doubts as to whether such a contextualist theory of relevance allowed us to give a properly reductive account of the concept of knowledge.

Our final position is as follows. Traditional naturalism maintains that it is possible to combine a view of philosophy as being constrained by science with a belief that epistemology can play a meliorative role in improving our epistemic practices. If this is tenable, it must be shown that the concept of knowledge is naturalistically acceptable: this is standardly taken to require a reductive definition of the concept of knowledge, and reliabilism is normally taken to offer the best chance of giving such an account. Our recent considerations indicate that the reliabilist strategy does not seem likely to be able to provide a reductive definition of knowledge. This poses serious problems for the project of traditional naturalism:

Process reliabilism and contextualism

the only prospect for success consistent with the fundamental aims of traditional naturalism seems to be to reconsider whether the concept of knowledge can figure unreduced in scientific theories. Failing that, traditional naturalism will have to be abandoned. However, what we have not seen is any reason to doubt whether a priori knowledge can be accommodated within the framework of traditional naturalism, assuming that that project can be sustained. For the purposes of the following, final chapter, we will assume that the project of traditional naturalism can be sustained, and that there is some naturalistically acceptable notion of knowledge available to serve as the basis of a definition of a priori knowledge. We will then look at some of the problems facing the development of an account of minimally experience-independent knowledge.

CHAPTER NINE

Nonexperientiality

Chapter five opened with a discussion of the options available to us in giving an account of experience-independence. Since setting out the options we have been arguing that it is inappropriate to characterise apriority by requiring unrevisability of a priori knowledge, or infallibility of a priori warrants but not of warrants in general. A priori knowledge and warrants do not differ in point of unrevisability or infallibility from knowledge and warrants in general.

Only the minimal notion of experience-independence remains undiscussed. On this account, a belief forming process is experience-independent just in case it has only nonexperiential inputs. A belief is warranted a priori just in case it is produced by an appropriate process which takes only nonexperiential inputs. The goal of this thesis is to defend the claim that there is no reason for traditional naturalists to be sceptical about a priori knowledge; this goal would have been served should we stop here, drawing only the negative conclusion that if we wish to defend naturalistic apriorism we had better be appropriately circumspect in our claims about the nature of a priori knowledge and a priori warrant.¹ But a more satisfying conclusion would go further and address some of the problems that must be faced by an account of a priori knowledge given in terms only of the minimal notion of experience-independence. Such is the goal of this final chapter. The discussion will necessarily be incomplete: while we can make some general observations about the reliabilist conception of nonexperientiality, a full treatment of the problem would require nothing less than giving a worked-out theory of minimal apriority (that is: a priori knowledge cashed out only in terms of the minimal notion of experience-independence). And the task of giving such a theory is daunting, even once the pitfalls associated with misguided theories of experience-independence are put behind us. I will discuss only two recent theories of the a

¹ Such circumspection would undermine Devitt's doubts about the possibility of naturalised apriorism (Devitt 1998), and also allow us to evade Kitcher's scepticism about apriorism in mathematics (Kitcher 1983 - discussed in chapters four and five of this thesis).

Nonexperientiality

priori; theories which, though they offer some interesting observations about apriorism naturalised, do not supply us with a full and working account of naturalistic a priori knowledge. Such an account has yet to be offered; the goal of this thesis will have been achieved if some of the obstacles in the way of an account have been identified and cleared aside.

Explaining "nonexperiential"

On the minimal notion, experience-independence is analysed in terms of nonexperientiality. On this account, a token belief forming process is experience-independent just in case it has only nonexperiential inputs.² Let us say that a (token) belief forming process which meets this condition is a nonexperiential belief forming process. The corresponding minimal notion of a priori warrant is: a belief is warranted a priori just in case it is produced by an appropriate nonexperiential belief forming process. A belief is known a priori just in case it is a priori warranted. The central question for a minimal, naturalist, theory of a priori knowledge is: what is it for an input to a process to be nonexperiential? Before addressing this, however, we should mention, if only to reject, a salient alternative in explaining "nonexperiential".

The alternative is to characterise a process as nonexperiential just in case no experiences were needed to produce that *process*: in the terminology of higher order processes³, this would be to require that the higher order belief forming process which produced the belief be itself nonexperiential. It is not at all clear whether this account will work: for one thing, the proposed analysis, *a belief forming process is*

² The sense of "nonexperiential" in which we are chiefly interested is a property of process tokens. This is primarily because we are interested in whether a token belief is warranted nonexperientially, and what warrants a belief is a token process. There is a sense of "nonexperiential" which applies to process types - for example, the process (type) of mental arithmetic is presumably nonexperiential - but by focusing on the level of types we would also be ignoring the possibility that tokens of *experiential* process types can produce a priori warranted beliefs. For example, vision is an experiential process, but we might think, with Kant, that certain instantiations of that process type can generate a priori warranted beliefs about geometry. I will not explicitly discuss such cases, but we should not rule them out by treating experientiality as a type-level property.

³ See chapters seven and eight.

Nonexperientiality

nonexperiential just in case it was brought about by a nonexperiential higher order process, has the term "nonexperiential" on both sides, and so does not have much potential to explain "nonexperiential". Further, it seems to commit us to an infinite regress of higher order nonexperiential processes. And, finally, there can be little doubt that, say, teaching can be effective in setting up processes which confer a priori warrant: one can be taught mathematical procedures which will enable one to arrive at a priori knowledge. But the process of being taught, which is here operating as a second-order belief forming process, is experiential. So it is not the case that a process is nonexperiential only if it was produced by a nonexperiential higher order process. We will not discuss this option further.

We return to the initial suggestion, on which nonexperientiality is linked to the status of the inputs to a process. This approach to explaining "nonexperientiality" is very natural, for it is standard to characterise processes by their typical inputs. So, the visual belief forming process is so-called because it takes in visual information; the memory belief forming process is so-called because it takes in stored memories. Standard practice, then, would support the view that we characterise a belief forming process as nonexperiential just in case it takes nonexperiential inputs. " α is a nonexperiential process" seems definable using a three clause recursive definition, as follows (for token belief forming process α , and belief p):

- (1) If α produces a belief in p on the basis of no inputs, then α is a nonexperiential process
- (2) If α produces a belief in p on the basis of only beliefs produced by a nonexperiential process then α is a nonexperiential process
- (3) Nothing else is a nonexperiential process.

Unfortunately, we need to say more than this, because there seem to be cases where nonexperiential belief forming processes just do take in beliefs produced by experiential processes in the process of producing nonexperientially warranted beliefs. For example: S buys seven apples and five pears, and concludes that she

Nonexperientiality

has bought twelve pieces of fruit; here it looks as if the a priori process of addition has taken as input some experiences and produced an experiential belief. But this will seem a problem only for as long as we refuse to separate out the different processes at work in the formation of S's conclusion; the task for the defender of naturalistic a priori knowledge is to indicate how the different processes might be distinguished. A different side to the same problem is set out forcefully by Goldman. Goldman reminds us that many of the processes we take to provide a priori knowledge, such as mathematical and geometrical reasoning, do require experiential inputs if they are to produce warranted belief.

Here is an argument for disputing the general knowability of logic without perception. Note first that in actual practice logicians learn logical truths by reading the proofs others have constructed, and constructing their own proofs with pencil and paper. These activities involve perception. The *a priorist* would contend, however, that perceptual reliance on inscriptions is in principle dispensable. One could, in principle, frame all the relevant representations in the mind. But is this true? Do human cognitive capacities enable one to construct, or follow, a complex proof in the mind with sufficient reliability to qualify for knowledge? That is doubtful. It is universally acknowledged that short-term, or working, memory is limited in capacity. When this capacity is exceeded, error rates increase sharply. It seems likely, therefore, that many phases (even single steps) of a truly complex proof will involve more material that can reliably be maintained in working memory. Therefore, a logical truth whose proof involves that degree of complexity can only be known with the help of external inscriptions, hence, only with the help of perception.⁴

Goldman concludes that there is no a priori knowledge. The challenge he poses for the apriorist is to give an account of nonexperiential processes which allows a certain role for experience in producing beliefs. This problem is, as just noted, closely related to the "apples and pears" problem given above. In the "apples and pears" case the experiential input triggers off the processes which result in S's belief about her total shopping; in the Goldman case of using visual aids to the processes, the experiences sustain the process as it runs. The common feature is that in both cases we need to show that the fact that the process seems to have experiential inputs does not render that process experiential.

⁴ Goldman 1986: 301-2.

Nonexperientiality

Goldman does allow that there are simple mathematical, logical and conceptual truths, beliefs in which can be warranted on the basis of wholly nonexperiential processes. One response to Goldman's challenge, therefore, would be to retreat on the scope of our claims to a priori knowledge: to limit the processes which we will consider a priori warrant-conferring to just those processes which clearly meet the definition of nonexperiential process given above. This, though, would be to concede too much. There is a strong intuition that the simple, undisputed cases of a priori knowledge have significant commonalities with complex, more problematic cases: the intuition can be expressed as follows: even in the complex cases where we have to use experiences to achieve a warranted belief, the experiences are in some sense incidental. This intuition lies at the core of our concept of a priori knowledge, and should not be given up lightly. Our problem, then, is to explain how it is that an experience can be incidental to the operation of a process. There seem to be two possibilities open to us.

(i) We could say that experience is incidental to the operation of a belief forming process just in case, although we have to use experience to gain warranted beliefs from the process, an idealised subject could gain warranted belief from that process without such use of experience. Our need for visual aids, etc., is put down to the limits of our powers of attention and visualisation; an idealised subject, it is proposed, has no such limits, and can know anything knowable a priori without such experiential aids. Experience, then, is incidental to the operation of a process α just in case an idealised subject could have gained warranted belief from α without experience.

But this appeal to an idealised subject is unconvincing. We should reflect that we don't really know what cognitive powers an idealised subject would have: we have no grasp of what an idealised subject could know a priori that is independent of our intuitions about our own a priori knowledge. Our assumption that an idealised subject could not know, say, physical laws a priori is based on the assumption that experience is *not* incidental to the process of coming to know such

Nonexperientiality

things.⁵ We assume that the idealised subject has a priori knowledge in just those areas where experience is incidental: but then the specification of the cognitive capacities of the idealised subject contains an appeal to the notion of the "incidentalness" of experience and cannot explain it. Further, even if an idealised subject could know *p* a priori, it would not obviously follow that we know it a priori. So it seems that the notion of an idealised subject cannot help us to understand what it is for an experience to be incidental.

(ii) A more promising way of explaining "incidentalness" would be via the suggestion that the experience of a diagram, a written note, a number of apples and pears, etc., in the a priori warranting of a belief, is just a *cue* for the initiation or sustaining of a properly nonexperiential process. As a first approximation, let us say, following Kim⁶, that an experience is a cue for a process just in case the having of that experience is causally relevant to the operation of the process, but not epistemically relevant to the status of the belief produced. Although the notion of a cue is widely adopted⁷, it so far merely reformulates the problem: to make progress, we need to know how an experience can be causally but not epistemically relevant to the formation of a belief. Understanding how this is possible seems especially difficult from the perspective of a naturalised epistemology which ties the epistemic status of a belief to its method of production. There might not, at first, seem to be room within such an account for a notion of a cue functioning in the way in which Kim requires.

Since we have defined "nonexperiential process" in terms of the inputs to processes, if we are to make sense of the idea of a nonexperiential process having an experiential cue we must not treat the cue as a genuine input: we need to find a way to allow cues to be causally relevant to the nonexperiential process without being

⁵ Goethe, for example, apparently thought that physical laws were knowable a priori through a process of visualisation (Bortoft 1996); to appeal to an idealised subject in disputing his claim would be to do no more than appeal to our intuition that such empirical facts cannot be known a priori. Incidentally, the claim that such empirical facts cannot be known a priori is cast in doubt by the possibility of innate a priori knowledge, which we will shortly discuss.

⁶ Kim 1981: 344, 350.

⁷ For example: Kant's notion of mathematical intuition, Chisholm's account of the role of experience in "intuitive induction" (Chisholm 1977: 39), and Kitcher's idea of a sufficient life (Kitcher 1983), are all ways of assigning experience the role of a cue.

Nonexperientiality

inputs to it. The distinction we need to appeal to seems to be that between a process and the casual background against which it takes place. We have already had occasion to appeal to this distinction, without commitment to a specific theory of causation which validates it. In chapter five we accepted that a bin fire in a building would not have caused a conflagration if the building had been flooded with water, without committing ourselves to saying that the non-floodedness of the building was a cause of the conflagration. However, we should also note that the non-floodedness of the building is causally connected in some way to the conflagration - indeed, assuming that all events have a common cause in the Big Bang, it seems that all events will be indirectly causally connected to each other. If this is right, then the non-floodedness of the building is causally connected to the conflagration without being a cause of the conflagration. This sort of distinction seems apt for explaining the notion of a cue. Tentatively, let us say that an event e is a cue for a process α just in case e is a prominent event in the near causal background to α , causally relevant to, but not an input to, α .

Whenever a specific event is claimed to be a cue there will be a temptation to say that it was in fact a genuine causal input; but it may be that disagreement here is really disagreement over the way the process in question should be individuated. For example, a tiny keystone is removed from a bridge, resulting in it collapsing, and crushing an object beneath it. If we focus on the crushing of the object itself, the removal of the keystone will appear as a cue: it is causally relevant to the crushing of the object but not an input into the process-of-crushing (the keystone itself did not actually crush the object, the bridge did). If we widen our perspective, the removal of the keystone may start to look like a cause of the crushing: perhaps the object was a police car, and Don Juan collapsed the bridge to thwart the police by crushing their car. Here we may want to say that the removal of the keystone was a cause of the crushing of the car, but we are now individuating the processes differently. Another example: the waving of a flag signals a driver of a train to leave the station. Again, if we focus on the process instantiated by the train's departure, we may say that the waving of the flag was a just a cue: it was the *driver* who caused the train to leave the station by pressing the accelerator. From a wider

Nonexperientiality

perspective, the flag may, say, be considered a cause of the process of the train leaving on time. So, whether an event counts as a cue or as a causal input into a process will depend on our criteria for appropriate individuation, but this does not undermine the idea of a cue: we need *some* way to individuate processes, and once we have settled on a way suited to our purposes, some events will count as cues, others as causes.

This notion of a cue helps us to understand how an experience can be causally relevant to the production of a belief without bearing on its epistemic status. The epistemic status of a belief is given by the recursive definition of a nonexperiential process given earlier. For a belief forming process to count as nonexperiential we will have to individuate it in such a way that all causally relevant experiences are counted as cues: they are relegated to the status of prominent conditions in the near causal background.

On this account, however, it might seem that it is largely a matter of convention whether we treat an experience as a cue rather than as genuinely significant to the epistemic status of the output belief. It might seem that an apriorist would seek an individuation of their candidates for a priori warranting processes which preserves their nonexperientiality, while an anti-apriorist would seek to show that input experiences are not merely cues, and that this conflict is incapable of meaningful resolution. However, while the account of cues just offered clearly does provide for fluidity in the way processes are individuated, it would be wrong to claim that whether there are any a priori warranting processes is purely a matter of convention. For a naturalised epistemologist, the individuation of belief forming processes will be a matter for cognitive psychology; whether cognitive psychology will find a place for a priori knowledge will depend primarily on the theory of minimal apriority which we adopt. Once a working theory has been found, the question of whether there is a priori knowledge will be decided by negotiation between that theory and cognitive psychology as a whole, and the outcome of this process of negotiation cannot be prejudged here. With Rey, I hold that the question of whether there is a priori knowledge is to an extent an a

Nonexperientiality

posteriori question.⁸ The apriorist may have strong intuitions that there is such knowledge, but should acknowledge the possibility that cognitive psychology will reveal that there is just no way to individuate belief forming processes which renders some of them nonexperiential. Philosophy contributes to this task by suggesting candidates for nonexperiential belief forming processes - that is, by offering theories of how a priori knowledge is possible. For the moderate naturalist, these theories must not appeal to concepts which are not scientifically acceptable, must not assume that a priori knowledge is unrevisable knowledge, and must not assume that a priori warrant is infallible in a way that warrant in general is not. I conclude this thesis by looking at two recent accounts of a priori knowledge which meet these requirements. The first I find to be unconvincing; the second I find to raise striking issues about the scope of a priori knowledge construed naturalistically. Neither of these accounts will receive anything like a full discussion: I raise them in order to highlight some issues that should be addressed by future discussions of naturalised a priori knowledge.

Boghossian's implicit definition account

In recent publications, Boghossian has advanced a theory which purports to explain our a priori knowledge of logic by rehabilitating the concept of analyticity.⁹ Hitherto, he suggests, philosophers have conflated two separate and independent notions of analyticity. The analytic explanation of necessity appeals, he suggests, to a "metaphysical" notion of analyticity, defined as follows:

Sentence *A* is (metaphysically) analytic iff *A*'s truth value is determined solely by its meaning, and not by the facts.

Boghossian argues strongly against the analytic theory of necessity, claiming that the theory is founded on a misunderstanding of the relationship between meaning and truth, but we shall not dwell on his arguments here. However, there is another

⁸ Rey 1998.

⁹ Boghossian 1994a, 1996, 1997.

Nonexperientiality

concept of analyticity, independent of the metaphysical concept, which Boghossian calls "epistemic analyticity". Roughly speaking, a sentence *A* which means *p* is epistemically analytic just in case: for *S* to grasp the meaning of *A* is sufficient for *S* to know *p*. Commitment to the existence of epistemically analytic sentences, Boghossian thinks, is a concomitant of rejecting the thesis of the indeterminacy of meaning¹⁰; hence, he maintains, philosophers cannot buy into Quine's rejection of analyticity (where analyticity is taken to be the epistemic notion) while maintaining that meaning is determinate. After defending the existence of epistemically analytic sentences in the face of Quinean doubts, Boghossian argues that logic is epistemically analytic; this is, he holds, because the axioms of logic are implicit definitions of the logical constants, and he upholds the notion of implicit definition by distinguishing it from conventionalism and non-factualism about logic. Then, the implicit definition account of the apriority of logic, in conjunction with the assumption that synonymy facts are knowable a priori, puts Boghossian in a position to explain other areas of a priori knowledge by appeal to Frege's account of analytic truth.

This is Boghossian's proposal in outline¹¹; but we need to look at it more closely.

The first problem with assessing Boghossian's account of a priori knowledge is that of getting a clear view of the notion of epistemic analyticity. Boghossian writes that a sentence *A* would be epistemically analytic if "mere grasp" of *A*'s meaning by *S* "sufficed" for *S* "being justified in holding" *A* true.¹² We shall take the idea of "grasping" a meaning as intuitively clear; the same, however, cannot be said for the idea of "being justified in holding *A* true"; nor, in view of Boghossian's later

¹⁰ Quine 1960 chapter two. There is an extended literature: see for example Miller 1998, Hale and Wright 1997, Kirk 1986, Hornstein 1982.

¹¹ It is a summary of his 1996 and 1997 papers.

¹² Boghossian's actual phrasing is as follows:

How could a factual statement *S* be known a priori by *T*, without the help of a special evidence-gathering faculty?

Here, it would seem, is one way: *If mere grasp of *S*'s meaning by *T* sufficed for *T*'s being justified in holding *S* true.*

Nonexperientiality

comments, for the idea of one condition being "sufficient" for another. We shall deal with the latter two problems in turn.

In a section devoted to epistemological preliminaries, Boghossian writes:

...we may say that for T to know that p is for T to justifiably hold S true [where S is a sentence which means p], with a strength sufficient for knowledge, and for S to be true. And to say that T knows p a priori is to say that T's warrant for holding S true is independent of outer, sensory experience.¹³

It is clear from this passage that Boghossian takes "warrant" and "justification" to be interchangeable. This being so, it is appropriate to ask to which single epistemic concept he intends these terms to refer. We can read Boghossian as implying that justification/warrant is that property which, had in enough quantity, turns true belief into knowledge; hence it seems that his justification/warrant is closer to our concept of warrant than the intuitive concept of justification. However, there are differences: Boghossian implies that his justification/warrant can be possessed by a belief, though not in sufficient quantity to turn that belief into knowledge. This makes it very different from the account adopted in previous chapters, according to which warrant is a matter of appropriate production: a property which, it seems plausible, does not admit of degrees. Boghossian also includes a truth clause in his definition of knowledge; and on our conception of warrant such a clause is otiose. However: the only place where the concept of justification/warrant enters Boghossian's account is in the definition of epistemic analyticity; and here it simply flags whatever positive epistemic status is afforded a proposition subsequent on its being grasped; moreover, since Boghossian is centrally interested in the epistemic status of beliefs in logical truths, which are necessarily true, we do not have to worry about what epistemic property such a belief retains if that belief turns false. So, as in the summary of Boghossian's account above, we can replace the reference to justification/warrant with a reference to knowledge. A sentence A which means

¹³ Boghossian 1997: 333.

Nonexperientiality

p will be epistemically analytic just in case: for S to grasp A 's meaning is sufficient for S to know p .

Next: sufficiency. Boghossian's formulation of the definition of epistemic analyticity suggests strongly that he takes grasp of the meaning of an epistemically analytic sentence to be a *sufficient condition* for knowledge of it. This would suggest that warrant for belief in the proposition expressed by an epistemically analytic sentence is gained simultaneously with it being grasped. This is a striking claim, and would clearly be epistemologically significant if it could be made out. However, Boghossian does not intend "sufficiency" to be read in this sense. For, as we will see, knowledge of what is expressed by an epistemically analytic sentence is arrived at by grasping the proposition expressed and making *inferences* based on it. Since we may refrain from drawing an inference¹⁴ - or even, through conceptual limitations, be unable to draw it - it cannot be that Boghossian intends "sufficient condition" by his appeal to sufficiency. That this is correct is clearly brought out by a reformulation of the notion of epistemic analyticity which Boghossian offers at the close of his 1997 paper.

My brief here has been to defend epistemic analyticity; and this requires showing only that certain sentences are such that, *if* a person knows the relevant facts about their meaning, *then* that person will be in a position to form a justified belief about their truth.¹⁵

The crucial change is the reference to being "in a position" to know. Since the inferences which mediate between our grasp of the meaning of the epistemically analytic sentence and our knowledge of it are meant to be a priori (as we will see), Boghossian must mean, by our being "in a position" to know, that we are in a position to know *without any additional experiential evidence*. Hence, fully spelled out, the definition of epistemic analyticity is as follows:

¹⁴ This point depends on our drawing a distinction between an inference and an implication, but that there is such a distinction should be uncontroversial (see Carroll 1895/1995).

¹⁵ Boghossian 1997: 357.

Nonexperientiality

Sentence *A* which means *p* is epistemically analytic iff: no more experience is required for *S* to know *p* than is required for *S* to grasp *p*.

This formulation does not, whereas Boghossian's initial formulation seems to, entail that one knows *p* if one grasps *p*. Again, this implication has to be removed because knowledge of *p* is arrived at through inferences which one might not make.¹⁶ This formulation also highlights the very close connection between epistemic analyticity and apriority considered as a property of a proposition. If no more experience is needed to know *p* than to grasp *p*, then *p* is an a priori proposition. Hence sentence *A* which means *p* is epistemically analytic just in case *A* expresses an a priori proposition. So Boghossian's claim to be able to explain areas of our a priori knowledge by appeal to the concept of epistemic analyticity is tendentious. Epistemic analyticity can explain a priori knowledge only in a solecistic sense. If Boghossian does make progress in explaining a priori knowledge, it must be in a different area of his account.

The potentially interesting part of his account concerns how a sentence comes to be epistemically analytic. In chapter one we distinguished between the topic and warrant conceptions of a priori knowledge; on the warrant conception, which we defended, the notion of an a priori warrant is considered fundamental and is used to define what it is for a proposition to be a priori. Given a definition of a priori warrant, we can define what it is for a proposition to be a priori by appeal to the following equivalence:

¹⁶ Boghossian might try to avoid the proposed modification to his account by claiming that, if *p* is expressed by an epistemically analytic sentence and is grasped by *S*, *S* knows *p* even where *S* does not draw the requisite inferences. This strategy is unlikely to work. First, since one can grasp a proposition without believing it, one may not even believe *p*; hence this strategy requires that we allow knowledge without belief. Examples of knowledge without belief are rarely convincing: see for example Goldman's discussions of the Chicken-Sexers and the Bloodsucking Tick in Stich 1975. (The question of the relationship between knowledge and belief is probably not best approached through considering isolated examples. Williamson 1995 treats the issue as one of conceptual priority, suggesting that the concept of knowledge is more fundamental than the concept of belief: belief, he suggests, might be "botched knowing".) Boghossian could respond by claiming that where *p* is expressed by an epistemically analytic sentence and is grasped, *p* is tacitly believed. An operationalist notion of belief, such that *S* believes *p* just in case *S* would assent to *p* if asked, might seem to support this: the "obviousness" of the epistemically analytic propositions make them such that *S* would assent to them if asked. However, such a notion of belief is very different from the one we actually have: suppose *S* believes some universal generalisation, to which there is an obscure counterexample *q*; then, on this account, *S* already believes *q*, since she would assent to it if asked (Tim Williamson suggested this example).

p is a priori iff $\exists S(S \text{ knows } p \text{ a priori})$.

As we have seen, Boghossian defines a priori knowledge through reference to the notion of an experience-independent warrant¹⁷: hence it seems he takes the notion of an a priori warrant to be fundamental. Thus, the claim that a sentence is epistemically analytic entails, via the second equivalence above, the claim that there is an experience-independent process available which can warrant beliefs in the proposition it expresses. What interests us, then, is the nature of this process.

Boghossian's account of a priori knowledge

Immediately before introducing the concept of epistemic analyticity, Boghossian refers to three classes of statements traditionally thought to be the objects of a priori knowledge.

...logical statements, exemplified by such truths as:

Either Brutus killed Caesar or he did not;

mathematical statements, such as:

$7 + 5 = 12$;

and conceptual truths, for instance:

All bachelors are unmarried.

It is not clear whether he takes the contents of these three classes to be exhaustive of the a priori knowable propositions, though, as we will shortly see, if he does think this then his account faces problems. In his discussion of epistemic analyticity, Boghossian makes a significant assumption: he assumes that any explanation of a

¹⁷ Boghossian 1997: 333. Quotation reproduced on page 215 of this chapter.

Nonexperientiality

priori knowledge must appeal to semantic properties of the sentence taken to be express an a priori proposition.

Turning, then, to the epistemological notion of analyticity, we immediately confront a serious puzzle: How could any sentence be analytic in this sense? How could mere grasp of a sentence's meaning justify someone in holding it true?

Clearly, the answer to this question has to be *semantical*: something about the sentence's meaning, or about the way that meaning is fixed, must explain how its truth is knowable in this special way.¹⁸

Boghossian considers two ways in which the way meaning is fixed could make a proposition a priori. These are: if the proposition is, as he puts it, "Frege-analytic"; or if it is an implicit definition. Boghossian does not consider whether there could be a third way in which semantic facts could be relevant to a priori knowledge; if there is a third way, it is certainly not clear what it could be; however, if there is no third way, Boghossian's account also faces difficulties, as we will see.

The first way in which the way in which meaning is fixed could be relevant to the epistemic status of a proposition is if that proposition is expressed by a sentence which is "Frege-analytic": that is, transformable into a logical truth through intersubstitution of synonyms.¹⁹ If both logical truths and synonymy relations are knowable a priori, propositions expressed by Frege-analytic sentences will be knowable a priori. If the antecedent of this conditional can be upheld, we will be able to explain our a priori knowledge of a large class of a priori propositions. Boghossian says little in defence of his assumption that synonymy relations will be so knowable, though notes that it is disputed.²⁰ Since the assumption is intuitively

¹⁸ *Ibid.*: 337.

¹⁹ The definition of "analytic" given by Frege in *Foundations of Arithmetic* (Frege 1950).

²⁰ Boghossian writes that on most externalist views of meaning, synonymy facts will not be knowable a priori (Boghossian 1997: 338): it is not clear that he is right about this. Certainly, causal and aetiological theories of intentional content have striking consequences for meaning and self-knowledge, but it is not clear that they force a rejection of the a priori knowability of synonymy facts. This issue is normally discussed under the heading of externalist self knowledge; much work has gone into showing that externalism is compatible with privileged self knowledge, see especially Burge 1988, Davidson 1987, Heil 1988, Wright 1994. This having been shown, the issue arises of whether externalism, in conjunction with the thesis of privileged self knowledge, allows too much a priori knowledge; whether it allows "privileged access to the world" (Sawyer 1998). The standard formulation of this problem is

plausible, we will pass over discussion of it. Clearly, though, an appeal to Frege-Analyticity is incapable of explaining our a priori knowledge of logic, because the assumption that Frege-analyticity can explain a priori knowledge depends on the assumption that logical truths are knowable a priori. Hence, if the appeal to Frege-analyticity is to be any help, we must provide an alternative explanation of our a priori knowledge of logic. As noted, Boghossian takes it that any explanation of a priori knowledge must be semantic, and so turns to the notion of implicit definition to give the desired account. Before doing so, he gives an extended argument that the adoption of a sceptical thesis about the existence of Frege-analyticity entails the adoption of a sceptical thesis about meaning itself. What he says is important and interesting; however, it is not directly relevant here, and therefore we should turn immediately to discussion of the core of his theory: the account of the apriority of logic in terms of implicit definition.

A sentence *A* is an implicit definition just in case *A* cannot express a falsehood without change of language; equivalently, the truth of what *A* expresses is constitutive of the meaning of a term that *A* involves.²¹ Boghossian notes that it seems impossible that logical constants could be defined explicitly: any such definition would presuppose that we had already defined them. Hence, an appeal to implicit definition seems to be required if we are to define logical constants. Boghossian endorses the following strategy:

It is by arbitrarily stipulating that certain sentences of logic are to be true, or that certain inferences are to be valid, that we attach a meaning to the logical constants. More specifically, a particular constant means that logical object, if any, which would make valid a specified set of sentences and/or inferences involving it.²²

the Brown-McKinsey argument, of which the first premise asserts that we have a priori self knowledge. Notably, and contrary to Boghossian's assumption that externalism entails failure of a priori knowledge of synonymy relations (which would, I am assuming, be closely connected with failure of a priori self knowledge), most authors accept the first premise of this argument. See, amongst others, Brown 1995, McKinsey 1991, Brueckner 1992, Boghossian 1997b, Miller 1997, Sawyer 1998. Discussion of the implications of the discussion of a priori knowledge offered in this thesis to the Brown-McKinsey argument will have to be left for future work.

²¹ Boghossian 1997: 348.

²² *Ibid.*

Nonexperientiality

If we adopt the implicit definition strategy as an account of the meanings of logical constants, then it seems clear how we will account for our a priori knowledge of logic. Boghossian suggests that the following sort of argument is in place:

- (1) If logical constant *C* is to mean what it does, then argument-form *A* has to be valid, for *C* means whatever logical object in fact makes *A* valid.
 - (2) *C* means what it does.
- Therefore,
- (3) *A* is valid.²³

Boghossian's setting up of this argument is not wholly satisfactory: the appeal to the subjunctive in the first premise is unclear, and the expression "*C* means what it does" is not a happy one. A more articulated version of the argument might run as follows, where "*A*" stands for some sentence expressing a logical truth, "*c*" for a logical constant:

- | | | |
|-----|---|----------------------------|
| (1) | <i>A</i> is an implicit definition of " <i>c</i> " | assumption |
| (2) | " <i>c</i> " is meaningful | assumption |
| (3) | If " <i>c</i> " means <i>c</i> then <i>A</i> is valid | from 1 |
| (4) | " <i>c</i> " means <i>c</i> | from 2 |
| (5) | <i>A</i> is valid | <i>modus ponens</i> , 3, 4 |

The argument certainly looks valid: though more work would have to be done to spell out the transitions from (1) to (3) and from (2) to (4), there is no reason to think that this couldn't be done successfully. The point of the argument, of course, is to show that (5) is known a priori. For (5) to be known a priori, assumptions (1) and (2) must also be known a priori; then, their a priori status will transfer to (5).²⁴ If reference to an argument of this sort can explain how we know logical truths a priori, then our a priori knowledge of propositions expressed by Frege-analytic sentences will have been demonstrated. Indeed, Boghossian also hints that he

²³ *Ibid.*

²⁴ To get a priori knowledge of (5) we have to follow the derivation, and so, as we have noted, simply grasping the meaning of an epistemically analytic sentence is not sufficient for knowing it to be true.

Nonexperientiality

believes that an argument of this form can be used to explain our a priori knowledge of classes of a priori propositions notoriously not assimilable to Frege analyticity: propositions such as "everything red is coloured", etc.²⁵

On Boghossian's account, then, reasoning according to this inference schema constitutes an appropriate belief forming process whereby logical truths come to be known. The process takes as inputs a priori warranted beliefs in (1) and (2), and consists of cognitive state transitions which produce and confer a priori warrant on a belief in (5). If this theory of a priori knowledge can be upheld, we will have been offered a description of a nonexperiential belief forming process - that is, we will have an account of minimal apriority. Unfortunately, it seems that Boghossian's theory can make at best a very limited contribution to our understanding of how a priori knowledge is possible, as the following remarks are intended to show.

As we have noted, for (5) to be known a priori, (1) and (2) must be known a priori; and yet it has not been shown that we can know (1) and (2) a priori. Boghossian, of course, recognises this, and as we have seen in a previous quotation²⁶ intends his account to be understood only as a conditional claim: in effect, that *if* knowledge of the premises can be shown to be a priori, *then* knowledge of the conclusion will also have been shown to be a priori. The interesting question, though, is whether there is a serious prospect of vindicating the assumption that (1) and (2) can be known a priori.

The issue of whether we know a priori that our words are meaningful, and hence whether (2) is known a priori, is controversial, and for the sake of argument we would do well to avoid these difficulties by accepting Boghossian's assumption that (2) is known a priori. It is consistent with this assumption that we do not know a priori *what* our words mean, how they got their meaning, or what role sentences expressing them play in our language. Hence accepting that we have a priori knowledge of (2) does not require that we accept that we have a priori knowledge of (1). Further, it does not seem obviously plausible to assume that (1) is known a

²⁵ *Ibid.*: 339.

²⁶ This chapter, page 216.

Nonexperientiality

priori. We have seen Boghossian suggest that our a priori knowledge is exhausted by three classes of truths: logical, mathematical and conceptual. (1) is clearly not a logical or mathematical truth; and conceptual truths are characteristically impossible to imagine to be false, while (1) has no such character. (1), in fact, appears to be a substantive empirical truth about language, and as such seems likely to be known only a posteriori. Boghossian passes by this problem, claiming that "there was never any real prospect of explaining apriority merely on the basis of a knowledge of propositional content."²⁷ However, this response is insufficient. The difficulty is: if we take seriously Boghossian's suggestion that a priori knowledge is restricted to knowledge of logical, mathematical and conceptual truths, then there is no way of explaining how (1) is known a priori. Perhaps, then, we should not take Boghossian's suggestion too seriously: but then the limitations of his project become clear. The explanation of a priori knowledge of logic through appeal to the concept of implicit definition presupposes the existence of a different type of a priori knowledge, about which Boghossian is silent. In the absence of an account, his theory tells us little that is substantive about nonexperientiality, or about a priori knowledge.²⁸

Innate and a priori knowledge

The final account of a priori knowledge we will consider draws from recent work by Carruthers.²⁹ Carruthers develops a reliabilist account of innate knowledge, and argues that innate knowledge is a priori knowledge. For our

²⁷ Boghossian 1997: 357.

²⁸ Horwich (Horwich 1997) argues that the procedure of implicit definition can be given a clear characterisation within the context of a use theory of meaning. On such a theory, our *deciding* to regard a sentence containing a term "c" as true commits us to a pattern of regularities regarding the use of the term "c", and constitutes a genuine implicit definition of "c". Horwich maintains that we can thereby achieve knowledge that certain sentences containing "c" are true, but that we have to consider this knowledge as vulnerable to future revision and so not a priori. In this thesis we have argued that the revisability of a belief *p* is not sufficient to show that *p* is not known a priori: hence it seems that Horwich's claim that his knowledge through implicit definition is not a priori knowledge can be challenged. Horwich may, then, have raised the possibility of a genuine explanation of a priori knowledge through the use of implicit definition, within the context of a theory of meaning as use; but this will have to be left for further work.

²⁹ Carruthers 1992.

Nonexperientiality

purposes, we do not need to consider whether any actually existing human knowledge is innate: we shall focus on Carruthers's arguments that innate knowledge is *possible*, and that innate knowledge is a priori knowledge. Carruthers's proposal, if it can be upheld, promises to have striking consequences for our understanding of a priori knowledge. For as we will see, it is arguable that if there can be innate knowledge at all, there can be innate knowledge of substantive propositions³⁰: thus, if innate knowledge is a priori knowledge then there can be substantive a priori knowledge. An account of a priori knowledge in terms of innate knowledge therefore has the potential to break sharply with our intuitions about apriority, and to engender a radical reconception of the notion of a priori knowledge. The intention here is to see whether Carruthers's account of innate knowledge can offer us an explanation of minimal a priori knowledge. First, we must determine what is his account of innate knowledge.

Carruthers assumes that belief is a necessary condition for knowledge, and that an account of innate knowledge would involve first giving an account of innateness of belief and then citing a belief forming process which can produce and warrant such beliefs.³¹ Accordingly, Carruthers's argument that there is innate a priori knowledge breaks naturally into three claims: (i) there are innate beliefs; (ii) there are belief forming processes which warrant innate beliefs; (iii) innate knowledge is a priori knowledge. We will discuss these in turn.

(i) *Innate beliefs.*

The idea of innate beliefs has always been controversial; nevertheless, there are good *prima facie* reasons for supposing that they are possible. The general idea of an innate trait is familiar from developmental biology, and the idea of an innate cognitive trait is relatively uncontroversial from ethology.^{32 33} Possessing a belief can

³⁰ I will use the term "substantive proposition" to refer to propositions which are contingent, do not essentially involve an indexical, and which concern specific, mind-independent features of the world.

³¹ The sort of approach is common. See for example essays by Stich, by Goldman and by Hart in Stich 1975, all of whom proceed by distinguishing the question of whether there are innate beliefs from the question of whether such beliefs can be warranted.

³² The classic work in this area is by Lorenz; see Schiller 1957.

³³ One reason to countenance innate cognitive traits derives from dissatisfaction with a purely empiricist account of concept acquisition. Carruthers argues that attempts to show that all concepts can be acquired from experience are unconvincing (Carruthers 1992: chapter 4.)

Nonexperientiality

be a trait; so perhaps some innate cognitive traits are innate beliefs.³⁴ In seeking to explain "innate belief" we should start by giving an account of the innateness of a trait generally.

It would not be a good idea to require of an innate trait that it be present from the birth of the organism that is supposed to possess it. For one thing, if we employ such a criterion the suggestion that there are innate beliefs would appear to be totally implausible, since most creatures seem to be born with very few cognitive traits: but we have just suggested that the idea of an innate belief is not obviously implausible.³⁵ Further, we have good independent reason to think that a trait can be innate despite only appearing later in the process of an organism's development: human pubic hair, and human language are salient examples. Both are clearly innate traits, but neither is present from birth. A *developmental* account of innateness allows that a trait might be innate in S and yet only make its appearance subsequent on certain classes of experience undergone by S; this notion of innateness forms the basis of Carruthers's account of innate knowledge.³⁶

Carruthers does not give an explicit account of developmental innateness, but suggests that a trait can be innate if it is "innately determined to make its appearance at some stage in childhood."³⁷ This suggests that we adopt a dispositional account of innateness, such that a trait is innate just in case an individual is disposed to manifest it in the course of normal development.³⁸ Perhaps, then, we should develop an account of an innate trait such that innateness is tied to the idea of development under normal conditions, as follows (for trait *t*):

t is (developmentally) innate in S iff S develops *t* in normal environments.

³⁴ Proposed cases of innate belief are often objected to on the grounds that the behavioural phenomena the appeal to innateness is meant to explain can be explained more parsimoniously by appeal to innate dispositions and reflexes rather than full-blown beliefs (see, for example, Lewis 1979, Harman 1974). But that individual cases can be disputed does not undermine the *prima facie* plausibility of the general argument that there can be innate beliefs.

³⁵ Compare one of Locke's arguments in the Essay, Book I, which is roughly as follows: an innate idea would command universal assent; but infants assent to no ideas; therefore there are no ideas which command universal assent; therefore there are no innate ideas. (Though couched in terms of innate ideas, it would be a simple matter to rewrite Locke's argument for innate beliefs.)

³⁶ Carruthers 1992: 51-2.

³⁷ *Ibid.*: 51.

³⁸ See Stich 1975, introduction, for an extended discussion of this idea.

Nonexperientiality

One natural way to employ this general definition of innateness in defining innateness of belief would simply be to substitute "a belief in p " for t . However, some have worried that the appeal to the concept of normality raises problems for such an account. Our definition of an innate trait leaves it underdetermined whether some traits are properly innate, as opposed to being acquired in every normal environment. Stich illustrates this underdetermination by reference to the difference between an innate disease and a susceptibility.

To suffer from an innate disease is to be disposed to acquire its symptoms at the characteristic time in the normal course of events. To be susceptible to a (non-innate) disease is to be disposed to acquire its symptoms under certain *special* circumstances. Certain toxic diseases, for example, can be acquired only by certain people. A susceptible person, when exposed to the toxic substance, will come down with the symptoms. At the extremes, the distinction seems clear enough. But notice how the two shade into each other. Suppose a person becomes ill after ingesting a certain amount of a particular chemical. Suppose also that the chemical occurs naturally in the drinking water of the person's community. Is this a case of an illness caused by the substance, or of an innate disease whose onset can be prevented by avoiding the substance? Vary the example, now, so that the substance is nitrogen in the air, and ask the same question.

These examples illustrate a central feature of the notion of an innate disease. There are commonly a host of necessary environmental conditions for the appearance of the symptoms of a disease. If these conditions all occur naturally or in the normal course of events, the symptoms will be counted as those of an innate disease. But it is often unclear whether the occurrence of a certain necessary condition *is* in the normal course of events. So it will often be unclear whether a person is afflicted with an innate disease or is, rather, susceptible to a (noninnate) disease.³⁹

Stich suggests that the problem of how to distinguish innate from ubiquitously acquired traits can be ignored with respect to innate physical traits, since we have a fairly good intuitive grasp of which conditions should be considered normal for physical development.⁴⁰ But he thinks that the problem cannot be ignored with respect to innate cognitive traits: for our intuitions about what *beliefs* arise in the

³⁹ Stich 1975: 7.

⁴⁰ *Ibid.*: 9.

Nonexperientiality

process of normal development "seem to swell the ranks of innate beliefs beyond all tolerable limits."⁴¹ To illustrate Stich's concern, consider the belief that *I have legs*: this is acquired in the course of normal development, but it doesn't seem to be innate.⁴² So simple substitution of "a belief in *p*" for *t* in the above definition will not do, for the unadorned definition will leave it too unclear which beliefs count as innate. Carruthers seems to concur, for he suggests two ways of giving a developmental account of innate belief, both of which would address Stich's problem:

[the first sense in which a] **belief might be innate** [is if] **it is acquired in any course of experience sufficient for forming beliefs at all[...]** The second sense in which an acquired belief might be innate would be if its existence were inexplicable on any model of learning, its content being such that it could not have been learned from the experiences that gave rise to it.⁴³

Carruthers seems to favour the second sense in which a belief might be developmentally innate, and by combining this with the general definition of developmental innateness given above we might seem to have a way of blocking the problem which Stich raises. Though Carruthers doesn't take an explicit stand on his definition of innate belief, it seems fair to interpret him as endorsing an account as follows:

Belief in *p* is innate in *S* iff *S* develops belief in *p* in normal environments and the existence of *S*'s belief in *p* is inexplicable on any model of learning.

Here, beliefs like "I have legs", which arise in the course of normal development but which could have been learned on the basis of experience, are blocked from being innate by the second conjunct. It is not necessary for our purposes that we decide whether any actual organisms entertain innate beliefs.⁴⁴

⁴¹ *Ibid.*

⁴² Jim Brown suggested this example.

⁴³ Carruthers 1992: 51-2. Carruthers's second sense corresponds to Stich's "input-output" model (see Stich 1975: ch.1).

⁴⁴ There is an extensive literature on this. Carruthers 1992 makes a strong case for the existence of innate beliefs about folk psychology. See especially Stich 1975, Harman 1974.

Nonexperientiality

This supplementation of the basic definition of developmental innateness is not wholly satisfactory. First, the extra conjunct seems to give a way of *recognising* whether a belief is innate, not give a criterion for it to *be* innate. Second, the extra conjunct strays dangerously close to tautologousness: given that "innate" and "learned" are generally taken to be correlative terms, to say that S's belief is inexplicable on any model of learning seems to be to say little more than that S's belief is innate. Finally, we might worry that the proposed definition is too strong. It does not seem obviously impossible, given that there are innate beliefs, that one should entertain a belief innately that one *could* have learned. Carruthers's reason for ruling out such beliefs is that by showing that a belief could have been learned one undercuts all reason for thinking it to be innate⁴⁵: but it does not follow immediately from this that the belief is *not* innate. To pursue this possibility we would need more details about the mechanism of innate belief generation; but although the definition of "innate belief" we are attributing to Carruthers is not wholly satisfactory, it is not wholly implausible either. We should assume that modifications could be made which would meet these worries, and turn to a more central aspect of the theory of innate knowledge.

(ii) Warrants for innate beliefs.

Even granted the possibility of innate beliefs, it is not obvious that there can be innate knowledge. Hart, for example, allows for innate beliefs while maintaining that the idea of innate knowledge is open to a serious objection.⁴⁶ Hart holds that justification (in the intuitive sense of the term) is necessary for knowledge. To possess a justification for a belief one needs to be able to give a reason for holding it; but innate beliefs, not having been acquired through the standard processes of learning, will seem to be held arbitrarily, and, he argues, appropriate justification will not be available for them. Hence they will not constitute knowledge.

Hart's objection has little force relative to the reliabilist conception of knowledge, whether the contextualist account proposed in this thesis, or the version

⁴⁵ Carruthers 1992: 53.

⁴⁶ Hart, in Stich 1975.

Nonexperientiality

outlined by Carruthers.⁴⁷ For a reliabilist, though lack of justification can sometimes undermine knowledge, possession of justification is not a necessary condition for possession of knowledge. The reliabilist will be willing to accept innate beliefs as knowledge provided that they were created by an appropriate belief forming process. Without rehearsing the arguments over the nature of warrant, we will focus on the requirement that appropriate belief forming processes must be reliable. Can a reliable process be identified which could generate innate beliefs? Carruthers thinks there can be.

Innate beliefs will count as known provided that the process through which they come to be innate is a reliable one (provided, that is, that the process tends to generate beliefs that are true). There are two possible candidates for such a process: divine intervention on the one hand, and evolution on the other. We could maintain, as most classical rationalists did, that innate beliefs are directly implanted in the mind by a veracious God. Or we could hold that innate beliefs have been acquired through evolution, via natural selection. Each of these process would most probably be reliable.⁴⁸

(Carruthers adds that a naturalist must reject an appeal to divine intervention, leaving only the process of natural selection.) The idea that evolution can create and warrant innate beliefs seems originally to have been made by Goldman, who writes:

In order for an innate belief to qualify as an item of knowledge, it must be causally related in an appropriate way to the fact to which it corresponds. But what might such a causal connection be like? What sort of causal process might this be? The answer, I suggest, is *evolutionary adaptation*. Suppose there is a general fact *p* about the environment of a certain animal species, or about the relationship between members of the species and their environment. This fact *p*, let us suppose, has great survival value for the members of the species. In particular, recognition or apprehension of this fact by a member of the species is a crucial factor in ensuring its survival. Under these conditions, it would not be

⁴⁷ Carruthers 1992: ch.5. Carruthers considers a parallel objection to the one here drawn from Hart, and rejects it on reliabilist grounds.

⁴⁸ Carruthers 1992: 77-8.

Nonexperientiality

surprising if, by a process of natural selection, the members of this species eventually come to be born with the belief that p.⁴⁹

If we can vindicate the idea that evolution can function as a reliable belief forming process we will be well on the way to showing that there can be innate knowledge. Not all the way there, since as we have seen, reliability is only a necessary condition for a process to confer warrant, not a sufficient one; however, it is such a central condition that to demonstrate that evolution is a reliable belief forming process would be to give considerable support to the idea of innate knowledge.

Carruthers (and Goldman, in the paper from which the above quotation comes) takes the notion of reliability involved in the notion of a warranting process to be a relatively uncomplicated notion, definable in terms of the statistical frequency of true beliefs. He suggests that natural selection would meet the requirement of reliability.

What reason is there for thinking that natural selection would be a reliable process, supposing that it resulted in some innate beliefs? Notice, first, that true belief has immense survival value for any organism, such as ourselves, much of whose behaviour is caused by the interaction of beliefs and desires. For in general an organism's projects will only succeed if based upon beliefs that are *true*. This is not to say, of course, that action undertaken on the basis of a true belief is guaranteed to succeed.[...] But often (though not always) the failure will result from the falsity of some other belief.⁵⁰

As Carruthers recognises, however, the set of beliefs which are valuable for survival is distinct from the set of true beliefs. Not all beliefs which are valuable for survival will be true (provided one stays in the appropriate countries, the false belief that all cars drive on the left of the road may be useful for survival⁵¹); nor are all true beliefs useful for survival (consider beliefs in certain moral imperatives - perhaps: one should always aid the weak against the strong). Natural selection will favour beliefs which are valuable for survival; to uphold the thesis that natural selection

⁴⁹ Goldman, in Stich 1975: 117.

⁵⁰ Carruthers 1992: 111-2. *Italics in original.*

⁵¹ Sterelny considers various examples of false but useful beliefs (Sterelny 1990: 131).

Nonexperientiality

construed as a belief forming process is reliable, therefore, we will have to give reason to think that, for the innate beliefs, there is a significant correlation between truth and survival-value.⁵² Carruthers finds such a reason in the observation that evolution is a process which operates over long periods of time.

[W]holly false beliefs will not have survival value in the long run, and in evolutionary selection it is the long run that matters. What seems undeniable is that organisms (of the sort that act on beliefs) will only survive, in general and in the long run, if they base their actions on beliefs that are true, or at least close to the truth. So if any innate beliefs have arisen through natural selection, we should expect them to be approximately true.⁵³

Clearly, a great deal more work needs to be done to determine whether evolution, or natural selection, favours truth over useful falsehoods. But for present purposes we should accept for the sake of argument that if evolution can be considered a belief forming process producing innate beliefs, then it will be reliable. Carruthers concludes that innate knowledge is possible; moreover, he claims that we actually do have innate knowledge of which sentences of our own language are well-formed, of the general spatiotemporal structure of the world, and of the basic structure of our own and others' psychology.⁵⁴ We should note, though, that although these are candidates for actual cases of innate knowledge in humans, the potential scope of innate knowledge is very much wider. For it seems very plausible that in a stable environment, fine-tuned adaptations will confer strong survival benefits: the more specific an innate belief, the more it could contribute to the survival of the organism. So it seems that Carruthers ought to accept that it is possible that there could be innate beliefs which encode very specific information about the believer's environment. Such beliefs could be in propositions which are contingent, do not essentially involve an indexical, and which concern specific, mind-independent features of the world - that is, they could be substantive. So it seems that the proponent of innate knowledge should accept that such knowledge is at least potentially of substantive propositions.

⁵² Lewis 1979 objects on these lines to the assigning of an epistemic role to evolution.

⁵³ Carruthers 1992: 112.

⁵⁴ *Ibid.*: ch.4, ch.8.

Nonexperientiality

(iii) *Innate knowledge is a priori knowledge.*

Carruthers takes it to be fairly obvious that innate knowledge is a priori knowledge.

But would innate knowledge, if it existed, at the same time be a priori? In one sense at least it would be. For, [...] one thing that can be meant by saying that something is known a priori is that our knowledge of it is independent of empirical support. In this sense, a priori knowledge is knowledge that has not been learned from experience, nor requiring support from experience to qualify as knowledge.⁵⁵

In explaining the notion of independence from empirical support by appealing to the notion of not being learned from experience, Carruthers clearly understands the notion of experience-independence in the sense of "nonexperientiality", and so on his accounts of innate knowledge and a priori knowledge, it seems that innate knowledge will indeed be a priori knowledge.⁵⁶

The striking consequence for the scope of a priori knowledge adverted to earlier stems from the conjunction of the conclusion to this argument with the point recently made that innate beliefs can encode specific information about the environment, and so be of substantive propositions. If evolution reliably produces innate belief in such a proposition, it will be known innately and hence known a priori. On Carruthers's account of innate knowledge, therefore, substantive a priori knowledge is possible.

The crucial question which Carruthers's account of innate a priori knowledge must face is whether innate knowledge is indeed nonexperiential knowledge. While we focus on the way beliefs arise in an individual, it does seem reasonable to say that innate beliefs are acquired nonexperientially, since they are not learned, but simply appear in the process of normal development. But the process of individual development is not the process being cited as the warrant for

⁵⁵ *Ibid.*: 78.

⁵⁶ Kitcher (1978) objects to this inference on the grounds that a priori knowledge is, while innate knowledge is not, unrevisable. We have already seen reason to reject appeals to the concept of unrevisability in characterising "experience-independence", and will not consider Kitcher's objection.

Nonexperientiality

innate beliefs: natural selection is meant to be playing this role. And natural selection is a process which transcends the life-history of any individual creature: the life-processes of individuals may be thought to compose part of the process of natural selection, but if it is natural selection which is being cited as a belief forming process, then the fact that S did not have to learn p to know p does not entail that S knows p on the basis of a nonexperiential process: what needs to be shown, rather, is that the process of natural selection itself is nonexperiential, and it is not at all clear whether this can be shown.

Put very roughly, the way in which a belief in p is meant to come to be an item of innate knowledge in S through natural selection seems to be as follows. At some point in the past, one of S's ancestors, call her A, undergoes a random genetic mutation which leads her, figuratively speaking, to "grow" a belief in p . That mutation is inherited by A's descendants, and hence S also comes to grow the belief during the course of her development. Thus far, this is a picture only of generation and transmission of an innate *belief*, not an innate warranted belief. It should be clear that S's innate belief is not warranted, since the process of random mutation which led to the formation of A's belief in p is not a reliable one. But, so far, the process of formation of S's belief in p does look wholly nonexperiential. The process whereby A comes to possess belief in p meets the requirements of our earlier definition of " α is a nonexperiential process", and there seems to be no reason to think that the process of inheritance whereby S comes to believe p will undermine this.⁵⁷

The picture so far, then, allows S to have an unwarranted, nonexperientially produced belief in p . The trouble is that in filling out the picture by explaining how S's belief in p is warranted, it is not at all clear that the feature of nonexperiential production is retained. The reason why S's belief in p is warranted is supposed to be that, in the circumstances, p enhances the fitness of the individuals who believe in it. Crudely speaking, we are to imagine that the rest of the population of which A is a member dies, without offspring, for lack of belief in p , leaving A alone to pass

⁵⁷ Environmental factors, such as the fact that S got the nourishment necessary for normal development, will be causally relevant to the process of S growing the belief; but if we have to treat these as "experiences" we should be able to relegate them to the status of causal "cues".

Nonexperientiality

on the genes for that belief to her descendants. But it is not totally counterintuitive to say that process whereby the non-*p*-believers came to die is in some sense experiential, since, lacking belief in *p*, they fail to cope with what the environment throws at them. It is just not clear whether we should count S's belief in *p* as warranted nonexperientially: A's own belief in *p* is created and handed on through nonexperiential processes, but the warrant for the belief is explained by the deaths of the non-*p*-believers. It's just not clear how we should individuate the processes in this scenario, and so it's not clear whether we should count S's belief in *p* as warranted nonexperientially. To clarify the situation we would need a much more accurate picture of the way natural selection works than that given by Carruthers or Goldman, and I am unable to provide such a picture.

However, there is another, more general reason to doubt that natural selection can play a role in generating nonexperientially warranted beliefs. This concerns the propriety of assigning natural selection a role in generating trait tokens. As we have already mentioned, warrant is a property of belief tokens: the explanation of how a belief acquires warrant calls for the specification of a process which explains the presence of that belief token. But it is at least doubtful that natural selection can be invoked to explain why an individual has the traits they actually have; rather, natural selection seems to provide an explanation of the frequency of traits in a population. Sober offers an example which illustrates this. Admission to a class requires that students be able to read at the third grade level. Selection for reading ability explains why the class comprises only students who read at the third grade level, but it does not explain, of any individual child, why that child reads at the third grade level. Analogously, natural selection can explain why all opossums have prehensile tails, but cannot explain why this or that opossum has a prehensile tail.⁵⁸ Hence, natural selection cannot explain why an individual has the innate beliefs that they do, and so cannot explain why those beliefs are warranted.

⁵⁸ Walsh 1998, discussing Sober 1984.

Nonexperientiality

These remarks concerning whether natural selection can be treated as a nonexperiential belief forming process are not intended to be conclusive; clearly, a great deal more work needs to be done on the issue. However, they suffice to show that Carruthers's account of innate knowledge is unconvincing as it stands, and cannot provide a firm basis for a naturalistic account of minimal a priori knowledge which appeals to the notion of innate knowledge. This is not to say, however, that an account of innate a priori knowledge cannot in principle be given: some account of the generation of adaptations must be offered, and it may be that a working account will substantiate the idea of innate a priori knowledge.⁵⁹

Conclusion

The goal of this thesis was to support the view that acceptance of naturalism is compatible with belief in the possibility of a priori knowledge. I claimed in chapter one that all arguments against this view are based on misconceptions of the

⁵⁹ When we focused on the way innate beliefs arose in an individual it seemed plausible to describe them as warranted nonexperientially; this plausibility was undermined by the appeal to natural selection to explain how these beliefs could be nonexperientially produced and warranted: not only is natural selection only dubiously nonexperiential, it seems incapable of explaining of any token belief how it came to be possessed by an individual. Perhaps a more plausible account of innate knowledge could be given by switching our focus back to the way the process of individual development generates innate beliefs. Ariew defines innateness by appeal to the developmental biological concept of *canalization*, where canalization is "the capacity to produce a particular definite end result in spite of a certain variability both in the initial situation from which development starts and in the conditions met with during its course" (Waddington 1975: 99, cited in Ariew 1996: S25). Ariew links the innateness of a trait to the degree of canalization of that trait: "the degree to which a biological trait is innate for a genotype is the degree to which a developmental pathway for individuals possessing that genotype is canalized." (*Ibid.*) An account of innateness in these terms would be stronger than the dispositional account: a canalized trait would manifest in all normal environments, but would also manifest in some environments that deviate from the norm. This gives it the potential to deal with Stich's worry that the dispositional account makes too many beliefs innate: the belief that I have legs would not be highly canalized, since there are environments in which I would not develop the belief. This account also links innateness to the process of development of an individual, so has the potential of explaining how an individual has the innate beliefs they have. Potentially, an account of innateness in terms of canalization could give us an explanation of how innate knowledge is possible. However, on Ariew's account there is no dichotomy between innate and learned traits: rather, innateness comes in degrees. If innate knowledge is characterised as a priori knowledge, this would seem to require a conception of apriority on which it too comes in degrees. Such an account would be strikingly different from the standard view, on which there is a strict dichotomy between a priori and a posteriori knowledge. The implications of Ariew's account of innateness for the theory of a priori knowledge would certainly repay further work.

Nonexperientiality

concepts of apriority and naturalism, but this was tendentious: we could just define "naturalism" and "apriority" such that a naturalistic a priori is ruled out; what I have been trying to show is that there is no reason to adopt such definitions.

I have taken it that the core notion of a priori knowledge is that of nonexperiential knowledge, and that the core claim of naturalism is that philosophy is not autonomous from science. The autonomy of philosophy, I have suggested, was bought at the cost of taking philosophy to be concerned only with analytic truths. The way to deny its autonomy, then, is to follow Quine in attacking the concept of analyticity, but the vision of naturalism that thereby arises looks very different depending on whether one also attacks the analytic theory of a priori knowledge itself. The chief misconception I have been concerned to remove is that the analytic theory is the only theory of a priori knowledge on offer.

While we retain the analytic theory, denying the autonomy of philosophy does indeed issue in scepticism about a priori knowledge. But if we reject that theory then the existence of a priori knowledge remains an open question: we will need to adduce substantive reasons to think that a priori knowledge cannot be naturalised (assuming, that is, that one does not follow Quine in the further step of dropping the normative dimension from epistemology *tout court*).⁶⁰ "Traditional" naturalist critics of the a priori such as Devitt and Kitcher realise this, for though they build clauses requiring scepticism about the a priori directly into their definitions of naturalism they also seek to defend the inclusion of these clauses on the basis of the rest of their naturalist theories. Thus, Kitcher argues for scepticism about the a priori on the basis of a reliabilist theory of experience-independent knowledge on which the standards for experience-independence are set very high, requiring unrevisability of the belief warranted independently of experience. Devitt also argues against a priori knowledge on the basis of unrevisability, and by claiming that the concept of apriority is obscure.

We've responded to these arguments by maintaining that no reason has been given in the first place for thinking that a priori knowledge is unrevisable

⁶⁰ I also think that dropping the analytic theory undercuts most of our grounds for thinking that science does not deal in a priori knowledge: if we give up the view that mathematics is a body of analytic truths it seems natural to describe mathematics as a science, and not just in a metaphorical sense.

Nonexperientiality

knowledge (apart from the sense in which all knowledge, being factive, is directly unrevisable, as the Kripke paradox shows) and by suggesting that the notion of a priori knowledge is no more obscure than the notion of a suitably individuated nonexperiential belief forming process. We haven't seen a theory of what these processes are, but the idea itself is not obviously unacceptable.

The project of providing a theory of minimalist a priori knowledge remains the chief issue left outstanding by this discussion. My own feeling is that it would be more profitable to pursue Carruthers's idea of innate a priori knowledge rather than Boghossian's attempt to reinstate analyticity. One of the striking consequences of a working account of innate knowledge might be a vindication of the rationalist idea of substantive a priori knowledge. This would be to break sharply with the assumptions about a priority handed down to us by recent empiricists, but looked at from a longer historical perspective may not necessarily be such a departure from our intuitions about the core aspects of the concept of a priori knowledge.

Bibliography

- ARIEW, A. 1996. "Innateness and Canalization", *Philosophy of Science*, 63 (proceedings), pp.S19-S27.
- ASPRAY, W. and KITCHER, P. (eds.) 1988. *History and Philosophy of Modern Mathematics: Minnesota Studies in the Philosophy of Science Vol. XI*. Minneapolis: University of Minnesota Press.
- AUDI, R. 1983. "Foundationalism, Epistemic Dependence, and Defeasibility", *Synthese*, 55, pp.119-39.
- AYER, A. J. 1936. *Language, Truth, and Logic*. London: Victor Gollancz.
1956. *The Problem of Knowledge*. Penguin.
- BAKER, G. P. and HACKER, P. M. S. 1985. *An Analytical Commentary on the Philosophical investigations Volume 2. Wittgenstein: Rules, Grammar and Necessity*. Oxford: Blackwell.
- BENACERRAF, P. 1973. "Mathematical Truth", *The Journal of Philosophy*, vol.LXX, no.19, pp.661-679.
- BIRD, A.J. 1998. *The Philosophy of Science*. London: UCL Press.
- BOGHOSSIAN, P.A. 1994. "Inferential Role Semantics and the Analytic/Synthetic Distinction", *Philosophical Studies*. Spring 1994, pp.109-122.
- 1994b. "The Transparency of Mental Content", *Philosophical Perspectives*, 8, pp.33-50.
1996. "Analyticity Reconsidered", *Nous* 30: 3, pp.360-391.
1997. "Analyticity", in Hale and Wright (eds.) 1997.
- 1997b. "What the Externalist Can Know A Priori", *Proceedings of the Aristotelian Association*, 97.
- BONJOUR, L. 1985. *The Structure of Empirical Knowledge*. Harvard University Press.
1992. "A Rationalist Manifesto", in Hanson and Hunter 1992.
1994. "Against Naturalised Epistemology", *MidWest Studies in Philosophy*, XIX, pp.283-300.
1998. *In Defense of Pure Reason: a Rationalist Account of A Priori Justification*. Cambridge: Cambridge University Press.
- BORTOFT, H. 1996. *The Wholeness of Nature: Goethe's Way Toward a Science of Conscious Participation in Nature*. New York: Lindisfarne.
- BOTTERILL, G. and CARRUTHERS, P. 1999. *Philosophy of Psychology*. Cambridge: Cambridge University Press.
- BOYD, R., GASPER, P. and TROUT, J.D. (eds) 1991. *The Philosophy of Science*. London: MIT Press.
- BROWN, J. 1995. "The Incompatibility of Anti-Individualism and Privileged Access", *Analysis* 55: 3, pp.149-156.

- BRUECKNER, A. 1992. "What an Anti-Individualist Knows *A Priori*", *Analysis* 52: 2, pp.111-118.
- BURNYEAT, M. 1990. *The Theaetetus of Plato*, (with a translation of Plato's Theaetetus by M.J. Levett rev. by Myles Burnyeat). Cambridge: Hackett.
- BURGE, T. 1979. "Individualism and the Mental", *MidWest Studies in Philosophy*, 4, pp.73-121.
1986. "Individualism and Psychology", *The Philosophical Review*, XCV, no.1, pp.3-45.
1988. "Individualism and Self-Knowledge", *The Journal of Philosophy* 85/11. pp.649-63. (Reprinted in Cassam 1994.)
1993. "Content Preservation", *The Philosophical Review*, vol.102, no.4, pp.457-488.
- BUTCHVAROV, P. 1970. *The Concept of Knowledge*. Evanston.
- BUTRICK, R. 1970. *Carnap on Meaning and Analyticity*. The Hague: Mouton and Co.
- BUTTERFIELD, J. 1986. *Language, Mind and Logic*. Cambridge University Press.
- CARNAP, R. 1937. *The Logical Syntax of Language*. London: Kegan Paul.
1956. *Meaning and Necessity*. Chicago: University of Chicago Press.
- CARROLL, L. 1895/1995. "What the Tortoise said to Achilles", *Mind* vol.IV, 1895; reprinted in *Mind* vol.104. 416, pp.691-694.
- CARRUTHERS, P. 1992. *Human Knowledge and Human Nature*. Oxford: Oxford University Press.
- CASSAM, Q. (ed.) 1994. *Self-Knowledge*. Oxford: Oxford University Press.
- CASULLO, A. 1988. "Revisability, Reliabilism and *A Priori* Knowledge", *Philosophy and Phenomenological Research*, vol.XLIX, no.2, pp.187- 213.
- CHISHOLM, R. 1977. *Theory of Knowledge* (2nd edition). Englewood Cliffs, N.J: Prentice-Hall.
- CHOMSKY, N. 1975. *Reflections on Language*. Pantheon.
1995. "Language and Nature", *Mind* vol.104. 413. pp.1-61.
- CLARK, M. 1963. "Knowledge and Grounds, a Reply to Mr. Gettier", *Analysis* 23.
- CLARKE, P. and HALE, R. (eds.) 1994. *Reading Putnam*. Oxford: Blackwell.
- CLARKE, A and CHALMERS, D. 1998. "The Extended Mind", *Analysis* 58: 1, pp.7-19.
- COFFA, A. 1993. *The Semantic Tradition from Kant to Carnap*. (ed. Wessels, L.) Cambridge: Cambridge University Press.
- COHEN, S. 1988. "How to be a Fallibilist", *Philosophical Perspectives* 2. pp.91-123. Atascadero, California: Ridgeview.
- CRAIG, E. 1990. *Knowledge and the State of Nature*. Oxford: Clarendon Press.
- CREATH, R. 1992. "Carnap's Conventionalism", *Synthese*, 93, pp.141-165.
- CROSSLEY, J.N. et.al. 1972. *What is Mathematical Logic?* Oxford: Oxford University Press.
- DANCY, J. (ed.) 1988. *Perceptual Knowledge*. Oxford: Oxford University Press.
1985. *An Introduction to Contemporary Epistemology*. Oxford: Blackwell.
- DANCY, J. and SOSA, E. (eds.) 1993. *A Companion to Epistemology*. Oxford: Blackwell Reference.
- DAVIDSON, D. 1987. "Knowing One's Own Mind", *Proceedings and Addresses of the American Philosophical Association*, 60, pp.441-58.

- DEROSE, K. 1995. "Solving the Sceptical Problem", *The Philosophical Review*, vol.104, no.1, pp.1-52.
- DESCARTES, R. 1986. *Meditations on First Philosophy; with Selections from the Objections and Replies* (Trans. Cottingham, J.). Cambridge: Cambridge University Press.
- DEVITT, M. 1998. "Naturalism and the A Priori", *Philosophical Studies* 92, nos.1-2, pp.45-65.
- DRETSKE, F. 1970. "Epistemic Operators", *Journal of Philosophy*, vol.LXVII, no.24, pp.1007-1023.
- DUMMETT, M. 1959. "Wittgenstein's Philosophy of Mathematics", *The Philosophical Review*, 68, pp.324-48.
1978. *Truth and Other Enigmas*. London: Duckworth.
- EDIDIN, A. 1984. "A Priori Knowledge for Fallibilists", *Philosophical Studies*, 46, pp.189-97.
- ETCHEMENDY, J. 1988. "Tarski on Truth and Logical Consequence", *Journal of Symbolic Logic*, vol.53, no.1, pp.51-79.
1990. *The Concept of Logical Consequence*. London: Harvard University Press.
- FIELD, H. 1998. "Epistemological Nonfactualism and the A Prioricity of Logic", *Philosophical Studies*, 92, pp.1-12.
- FODOR, J. 1975. *The Language of Thought*. New York: Crowell.
1987. *Psychosemantics*. London: MIT Press.
- FODOR, J. and LEPORE, E. 1992. *Holism: A Shopper's Guide*. Oxford: Blackwell.
- FOGELIN, R. J. 1994. *Pyrrhonian Reflections on Knowledge and Justification*. Oxford: Oxford University Press.
- FOLEY, R. 1994. "Quine and Naturalised Epistemology", *MidWest Studies in Philosophy*, XIX, pp.243-260.
- FREGE, G. 1950. *Foundations of Arithmetic*, (trans. Austin J.L.), Oxford: Blackwell.
- GETTIER, E. 1963. "Is Justified True Belief Knowledge?", *Analysis* 23, pp.121-123.
- GIAQUINTO, M. 1996. "Non-Analytic Conceptual Knowledge", *Mind*, vol. 105. 418, pp.249-267.
- GOLDMAN, A. 1967. "A Causal Theory of Knowing", *The Journal of Philosophy*, vol. LXIV, No. 12, pp.357-372.
1976. "Discrimination and Perceptual Knowledge", *The Journal of Philosophy*, vol. LXXIII, no.20, pp.771-791.
1986. *Epistemology and Cognition*. London: Harvard University Press.
1988. "Strong and Weak Justification", *Philosophical Perspectives* 2, *Epistemology* 1988, Tomberlin (ed.), pp.51-3.
1992. "What is Justified Belief?", in Kornblith (ed.) 1992 (first published in *Justification and Knowledge* (Reidal 1979)).
1994. "Naturalistic Epistemology and Reliabilism", *MidWest Studies in Philosophy*, XIX, pp.301-320.
- GRAVES, C., KATZ, J. et al. 1973. "Tacit Knowledge", *The Journal of Philosophy*, 70, pp.318-31.

- GRICE, P. 1961. "A Causal Theory of Perception", *Proceedings of the Aristotelian Society* supp. vol. xxxv. (reprinted in Dancy 1988).
1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- GRICE, P. and STRAWSON, P.F. 1956. "In Defense of a Dogma", *The Philosophical Review*, 65, pp.141-158.
- GUTTENPLAN, S. 1975. *Mind and Language*. Oxford: Oxford University Press.
- HALE, B. 1987. *Abstract Objects*. Oxford: Blackwell.
1999. "On Some Arguments for the Necessity of Necessity", *Mind*, vol.108. 429, pp.23-52.
- HALE, B. and WRIGHT, C. (eds.) 1997. *A Companion to the Philosophy of Language*. Oxford: Blackwell.
- HANSON, P and HUNTER, B. 1992. *The Return of the A Priori: Canadian Journal of Philosophy Supplement*. University of Calgary Press.
- HARMAN, G. 1967-8. "Quine on Meaning and Existence I," *Review of Metaphysics* 21.
1973. *Thought*. Princeton: Princeton University Press.
1974(ed.). *On Noam Chomsky: Critical Essays*. Garden City, New York: Anchor Press.
1996. "Analyticity Regained", *Nous* 30: 3, pp.392-400.
- HART, W.D. 1975. "Innate Ideas and A Priori Knowledge". In Stich 1975, pp.107-108.
- HEIL, J. 1988. "Privileged Access", *Mind* vol. 97. 386, pp.238-251.
- HONDERICH, T. 1995. *Oxford Companion to Philosophy*. Oxford University Press.
- HOOKWAY, C. 1988. *Quine: Language, Experience and Reality*. Cambridge: Polity.
- HORNSTEIN, N. 1982. "Foundationalism and Quine's Indeterminacy of Translation Thesis", *Social Research* 49, pp.32-67.
- HORWICH, P. 1997. 'Implicit Definition, Analytic Truth, and Apriori Knowledge', *Nous* 31: 4, pp.423-440.
- HUGHES, G. E. and CRESSWELL, M. J. 1968. *A New Introduction to Modal Logic*. New York: Routledge.
- HUNTER, G. 1995. "Quine's "Two Dogmas of Empiricism": or The Power of Bad Logic". *Philosophical Investigations* 18: 4.
1996. *Metalogic : an Introduction to the Metatheory of Standard First Order Logic*. London: University of California Press.
- HYLTON, P. 1994. "Quine's Naturalism", *MidWest Studies in Philosophy*, XIX, pp.261-282.
- KANT, I. 1934. *Immanuel Kant's Critique of Pure Reason*. Trans. Norman Kemp Smith. London: Macmillan.
- KAPLAN, M. 1994. "Epistemology Denatured", *MidWest Studies in Philosophy*, XIX, pp.350-365.
- KIM, J. 1981. "The Role of Perception in A Priori Knowledge: Some Remarks", *Philosophical Studies* 40, pp.339-354.

1988. "What is "Naturalised Epistemology"?", *Philosophical Perspectives*, 2, *Epistemology*, (Tomberlin, J. (ed.)).
1993. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.
- KIRK, R. 1986. *Translation Determined*. Oxford: Clarendon.
- KITCHER, P. 1978. "The Nativist's Dilemma", *The Philosophical Quarterly*, vol.28, no.110, pp.1-16.
1980. "A Priori Knowledge", *The Philosophical Review*, vol.80, no.1, pp.3-23.
- 1980b. "Apriority and Necessity", *Australasian Journal of Philosophy*, 58, pp.89-101.
1983. *The Nature of Mathematical Knowledge*. Oxford: Oxford University Press.
1992. "The Naturalists Return", *The Philosophical Review*, vol.101, no.1, pp.53-113.
- KORNBLITH, H. (ed.) 1992. *Naturalizing Epistemology*. London: MIT Press.
1994. "Naturalism: Both Metaphysical and Epistemological", *MidWest Studies in Philosophy*, XIX, pp.39-52.
- KRIPKE, S. 1980. *Naming and Necessity*. Blackwell. (Originally published in *Semantics of Natural Language*. Davidson, D. and Harman, G. (eds). Dordrecht: Reidel, 1972.)
- LEWIS, D. 1973. "Causation", *Journal of Philosophy*, 70, pp.556-67. Reprinted in Sosa and Tooley 1993, ch.XI.
- 1973b. *Counterfactuals*. Oxford: Blackwell.
1996. "Elusive Knowledge", *Australasian Journal of Philosophy*, vol. 74, no.4, pp.549-567.
- LEWIS, H. 1979. "The Argument from Evolution", *Proceedings of the Aristotelian Society Supplementary Volume* 53, pp. 207-21.
- LIPTON, P. 1991. *Inference to the Best Explanation*. London: Routledge.
- LOCKE, J. 1975. *Essay Concerning Human Understanding*. Nidditch, P. H. (ed.). Oxford: Clarendon.
- LORENZ, K. 1957. "The Nature of Instincts", in Schiller 1957, pp.129-175.
- MCCULLOCH, G. 1989. *The Game of the Name : Introducing Logic, Language, and Mind*. Oxford: Clarendon.
- MCFETRIDGE, I. G. 1990. *Logical Necessity and Other Essays* (Haldane, J. and Scruton, R. eds.). London: Aristotelian Society.
- MCKINSEY, M. 1991. "Anti-Individualism and Privileged Access", *Analysis* 51, pp.9-16.
- MELLOR, D.H. 1991. *Matters of Metaphysics*. Cambridge: Cambridge University Press.
- MENDELSON, E. 1964. *Introduction to Mathematical Logic*. D. Van Nostrand.
- MILL, J.S. 1868. *A System of Logic, Ratiocinative and Inductive*. London: Longmans.
- MILLER, A. 1998. *Philosophy of Language*. London: UCL Press.
- MILLER, R.W. 1997. "Externalist Self-Knowledge and the Scope of the A Priori", *Analysis* 57: 1, pp.67-75.
- MOORE, A.W. 1993. *Meaning and Reference*. Oxford: Oxford University Press.
- MOSER, P. 1987. *A Priori Knowledge*. Oxford: Oxford University Press.

- NOZICK, R. 1981. *Philosophical Explanations*. Oxford: Clarendon.
- O'LEARY-HAWTHORNE, J. (ed). 1998. *A Priori Knowledge: Philosophical Studies*, vol. 92, nos. 1-2.
- PAPINEAU, D. 1993. *Philosophical Naturalism*. Oxford: Blackwell.
1996. *The Philosophy of Science*. Oxford: Oxford University Press.
- PEACOCKE, C. 1992. *A Study of Concepts*. Cambridge, Mass: MIT Press.
1993. "How Are A Priori Truths Possible?", *European Journal of Philosophy*, vol. 1, pp.175-199.
1999. *Being Known*. Oxford: Oxford University Press.
- PERRY, J. 1986. "Thought Without Representation", *Proceedings of the Aristotelian Society*, Supplementary Volume 60, pp.137-151.
- PLANTINGA, A. 1993. *Warrant: The Current Debate*. Oxford: Oxford University Press.
- POLLOCK, J. L. 1986. *Contemporary Theories of Knowledge*. Totowa, N.J.: Rowman & Littlefield.
- PRIEST, G. 1979. "Two Dogmas of Quineanism", *The Philosophical Quarterly*, vol.29, no.117, pp.289-301.
- PRIOR, A. 1960. "The Runabout Inference-Ticket", *Analysis* 21: 2, pp.38-39.
- PUTNAM, H. 1975. *Mind, Language and Reality*. Cambridge: Cambridge University Press.
1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.
1983. *Realism and reason: Philosophical Papers Volume 3*. Cambridge: Cambridge University Press.
- QUINE, W.V.O. 1936. "Truth by Convention", *Philosophical Essays for A.N. Whitehead*. New York: Longmans. Reprinted in Quine 1966.
1951. "Two Dogmas of Empiricism", *Philosophical Review*, vol.LX pp.20-43. Reprinted in Quine 1953.
1953. *From a Logical Point of View: 9 Logico-Philosophical Essays*. Cambridge, Mass.: Harvard University Press.
1960. *Word and Object*. Cambridge: Technology Press of the Massachusetts Institute of Technology.
1966. *Ways of Paradox and Other Essays*. New York: Random House.
1969. *Ontological Relativity and Other Essays*. New York: Columbia University Press.
1970. *Philosophy of Logic*. Englewood Cliffs, N.J.: Prentice-Hall.
1970. "On the Reasons for Indeterminacy of Translation", *Journal of Philosophy* 67, pp.178-83.
1991. "Two Dogmas in Retrospect", *Canadian Journal of Philosophy*, 21, pp.265-274.
- READ, S. 1993. "The Unavailability of What We Mean I: a Reply to Quine and Fodor and LePore". *Gazier Philosopher Studien*, 46, pp.61-101.
1994. "Formal and Material Consequence", *Journal of Philosophical Logic* 23, pp.247-65.
1995. *Thinking about Logic: an Introduction to the Philosophy of Logic*. Oxford: Oxford University Press.

- RESCHER, N. 1984. *The Limits of Science*. University of California Press.
- REY, G. 1998. "A Naturalistic A Priori", *Philosophical Studies*, 92, nos.1-2, pp.25-43.
- RORTY, R. 1980. *Philosophy and the Mirror of Nature*. Oxford: Blackwell.
- RUBEN, D.H. 1993. *Explanation*. Oxford: Oxford University Press.
- RUSSELL, B. 1912. *The Problems of Philosophy*. Oxford University Press.
- SALMON, W. 1967. *The Foundations of Scientific Inference*. Pittsburgh.
1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
1998. *Causality and Explanation*. Oxford: Oxford University Press.
- SAWYER, S. 1998. "Privileged Access to the World", *Australasian Journal of Philosophy*, vol. 76, no. 4, pp.523-533.
- SCHIFFER, S. 1996. "Contextualist Solutions to Scepticism", *Proceedings of the Aristotelian Society*, vol.XCVI, pp.317-333.
- SCHILLER, C.H. (ed.) 1957. *Instinctive Behaviour*. New York: International University Press.
- SELLARS, W. 1963. *Science, Perception and Reality*. London: Routledge and Kegan Paul.
- SOBER, E. 1981. "Revisability, A Priori Truth, and Evolution", *Australasian Journal of Philosophy*, vol.59, no.1.
1984. *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. London: MIT.
1993. *Philosophy of Biology*. Oxford : Oxford University Press.
- SOSA, E. and TOOLEY, M. (eds). 1993. *Causation: A Reader*. Oxford: Oxford University Press.
- STERELNY, K. 1990. *The Representational Theory of Mind*. Oxford: Basil Blackwell.
- STICH, S. 1975. *Innate Ideas*. University of California.
- STINE, G.C. 1976. "Skepticism, Relevant Alternatives, and Deductive Closure", *Philosophical Studies*, 26, pp.249-261.
- SUMMERFIELD, D. 1991. "Modest A Priori Knowledge", *Philosophy and Phenomenological Research*, vol. LI, no.1, pp.38-66.
- TARSKI, A. 1956. *Logic, Semantics, Metamathematics: papers from 1923 to 1938*. Trans. J.H. Woodger. Oxford: Clarendon Press.
- TIDMAN, P. 1996. "The Justification of A Priori Intuitions", *Philosophy and Phenomenological Research*, vol. LXI, no.1, pp.161-171.
- UNGER, P. 1975. *Ignorance: a Case for Scepticism*. Oxford: Oxford University Press.
- WADDINGTON, C.H. 1975. *The Evolution of an Evolutionist*. Ithaca: Cornell University Press, pp.36-59.
- WALSH, D.M. 1998. "The Scope of Selection: Sober and Neander on What Natural Selection Explains", *Australasian Journal of Philosophy*, vol. 76, no.2, pp.250-264.
2000. "Chasing Shadows: Natural Selection and Adaptation". *Studies in the History and Philosophy of Biology and Biomedical Science*, vol. 31, no.1, pp.135-153.

- WILLIAMSON, T. 1986. "The Contingent *A Priori*: has it anything to do with indexicals?", *Analysis* 46: 3, pp.113-117.
1994. *Vagueness*. Routledge.
1995. "Is Knowing a State of Mind?", *Mind*, vol. 104. 415, pp.535-565.
- 1995b. "Does Assertibility Satisfy the S4 Axiom?" *Critica*, vol.XXVII, no.81, pp.3-26.
- 1996a. "Cognitive Homelessness", *The Journal of Philosophy*, 93. pp.554-573.
- 1996b. "Knowing and Asserting", *The Philosophical Review*, 105, pp.489-523.
1997. "Knowledge as Evidence", *Mind*, vol. 106. 424, pp.717-741.
- 1998a. "Conditionalizing on Knowledge", *The British Journal for the Philosophy of Science* 49, 1, pp. 89-121.
- 1998b. "The Broadness of the Mental: Some Logical Considerations" in J. Tomberlin (ed.) *Philosophical Perspectives 12: Language, Mind and Ontology*, pp.389-410.
- WITTGENSTEIN, L. 1953. *Philosophical Investigations*. Oxford: Blackwell.
1961. *Tractatus Logico-Philosophicus*. (Trans. D. F. Pears and B. F. McGuinness). London: Routledge Kegan Paul.
1974. *Philosophical grammar*, edited by Rush Rhees, trans. Anthony Kenny. Oxford: Blackwell.
- WRIGHT, C. 1986. "Inventing Logical Necessity" in Butterfield 1986.
1991. "Scepticism and Dreaming: Imploding the Demon", *Mind*, vol. 100. 397, pp.87-116.
1994. "On Putnam's Proof that We Are Not Brains-in-a-Vat", *Reading Putnam*, ed. P. Clark and R. Hale. Blackwell.
1997. "The Indeterminacy of Translation", *A Companion to the Philosophy of Language* (ed. Hale and Wright). Blackwell.
- ZAGZEBSKI, L. 1994. "The Inescapability of Gettier Problems", *The Philosophical Quarterly*, vol.44, no.174, pp.65-73.